ARMY RESEARCH OFFICE

PROCEEDINGS
OF THE
1962 ARMY SCIENCE CONFERENCE
UNITED STATES MILITARY ACADEMY, WEST POINT, N. Y.
20-22 JUNE 1962

VOLUME II
(Principal Authors "L" thru "Z")

OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

☆ ☆ ☆ ☆ ☆ ☆ ☆

HEADQUARTERS
DEPARTMENT OF THE ARMY
Washington 25, D.C.

AVAILABILITY NOTICE

Qualified requestors may obtain copies of individual papers contained in this report from ASTIA. Requests for complete volumes should be sent to the address in the following paragraph.

Government agencies not authorized to utilize ASTIA may obtain copies of individual papers or the complete report from the Office, Chief of Research and Development, Department of the Army, Washington 25, D. C., Attention: Research Support Division.

Non - government agencies not accredited to ASTIA obtain copies of individual papers from the Office of Technical Services, Department of Commerce, Washington 25, D. C.

HEADQUARTERS
DEPARTMENT OF THE ARMY
OFFICE OF THE ADJUTANT GENERAL
WASHINGTON 25, D. C.

IN REPLY REFER TO
AGAM-P (M) 337 (17 Aug 62) CRD/P

28 August 1962

SUBJECT: Proceedings of Army Science Conference, 1962

TO: See Distribution

1. Inclosed for your information and use is Volume Two of the "Proceedings of Army Science Conference, 1962," which is comprised of the unclassified papers which were read at the conference.

2. Volumes One and Three of the Proceedings will be distributed in the near future.

By Order of the Secretary of the Army:

J. C. Lambert

J. C. LAMBERT
Major General, USA
The Adjutant General

1 Incl
Vol II, Proceedings of Army Science Conf, 1962

DISTRIBUTION:
As Indicated in Distr List in Volume

P R O C E E D I N G S

O F   T H E

1962 ARMY SCIENCE CONFERENCE

UNITED STATES MILITARY ACADEMY, WEST POINT, N. Y.

20-22 JUNE 1962

V O L U M E   II

(Principal Authors "L" thru "Z")

INTRODUCTION

1. The Proceedings of the 1962 Army Science Conference are a compilation of all papers delivered at the conference. Its purposes are to:

   a. Provide a historical record of the conference.

   b. Serve as source material for scientists engaged in research for the Army.

   c. Enable conference participants to review papers in which they have a special interest.

2. The report consists of three volumes, as follows:

   Volume 1 - Unclassified papers
   Volume 2 - Unclassified papers
   Volume 3 - Classified papers

Dwight E. Beach

DWIGHT E. BEACH
Lieutenant General, GS
Chief of Research and Development

TABLE OF CONTENTS

PROCEEDINGS OF 1962 ARMY SCIENCE CONFERENCE

PROBLEMS AND PROGRESS IN THE STUDY OF ORAL TOXICITY OF BACTERIAL TOXINS

CARL LAMANNA
LIFE SCIENCES DIVISION, ARMY RESEARCH OFFICE
WASHINGTON 25, D.C.

In common parlance, the term food poisoning is associated with the consumption of food in which harmful bacteria have grown and produced products which, when ingested, are upsetting to the unsuspecting victim. Caveat consumere - let the consumer beware. A large number of different and phylogenetically unrelated bacteria have been implicated in food poisoning, some frequently, and others so infrequently that their capacity to cause food poisoning remains suspect.

The signs of food poisoning should be traceable to the effects of particular compounds. In other words, one aspect of basic scientific investigations of food poisoning should be the isolation and characterization of specific compounds which, by their biological effects, can account for the signs and pathology of food poisoning. It is surprising that the number of such compounds from bacteria which have been so specifically identified is small. Most interesting is that the few such materials which are characterizable as to their general nature are proteins. As such, they are properly classifiable as exotoxins, since they appear to be secreted, excreted, or otherwise released into the medium in which the bacteria grow, are antigenic, and can be neutralized by specific antibody. These are properties traditionally associated by the bacteriologist with the substances he calls exotoxins.

Scientific nomenclature contributes to clear thinking by insistence upon the use of precise definitions to identify objects and phenomena. It is in this spirit that I propose the limitation of the term enterotoxin to a certain kind of oral poison. When an exotoxin is swallowed, its harmful effect can be due to direct action on the tissue of the alimentary tract, in which case the toxin is properly spoken of as an enterotoxin: one which by direct contact specifically affects the behavior of intestinal cells. The other possibility is that the toxin does not act directly on the gut, but

rather is absorbed from the alimentary tract and acts specifically at sites remote from the intestinal lumen. In this situation, we can speak of the toxin as being an oral poison, since it causes harm when swallowed, but should not speak of it as an enterotoxin. Any effects observed on the alimentary tract would be the secondary consequences of action at some distant primary site: an extra-alimentary or extra-lumen site. Such a situation poses an interesting critical problem missing for the case of a true enterotoxin, namely, how can a protein escape the intestinal barriers to the absorption of large-sized molecules?

If a protein can act as an oral poison, we are inclined to make certain common sense inferences as to the properties that are responsible for its being an oral poison. We may reason that either the molecule as a whole, or some smaller specific piece or toxophore, must have some special resistance to the intestinal environment which is harmful to the maintenance of the structural integrity of a protein, for example the actions of proteolytic enzymes. If the toxin can act at a distance from the intestinal tract, we may also be led to suppose that the toxin must possess special properties to account for its transport across the intestinal permeability barriers. It is the validity of these two inferences, when tested against facts, that I will discuss, using botulinal toxin as the model of an orally poisonous exotoxin, and one which has its primary action remote from the digestive system. When I can, I shall consider the biochemical problems encountered. The goals of biochemistry are to relate toxicity to the chemical structure of the toxin molecule and to identify those factors of susceptibility of the host to the toxin which have a biochemical basis.

In botulism, the harm done to the animal victim is the result of a toxemia following ingestion of poisoned food and does not involve an infectious process. Thus we are not involved with considerations of the inflammatory process which complicate problems of toxin absorption and action.

Botulinal toxins appear to be simple proteins. It is not possible to say how the property of toxicity arises in this protein. The search for localized toxophoric groups within the structure of the protein molecule has been negative to date. A variety of unrelated physical agents and chemical reactions cause loss of toxicity. A recent hope that fluorescence of the toxin at 3300 Å after activation at 2900 Å is associated with toxicity has not been substantiated, since detoxification, for example by urea, can be accomplished without an accompanying loss of fluorescence (1). The available facts do not prove, but do support, a hypothesis that attributes toxicity to the maintenance of the structural integrity of the protein molecule as a whole.

Being willing to accept the guess that the toxic structure of botulinal toxin is a complete protein molecule raises the

question of how such a protein can escape detoxification in running the gamut of digestive juices. While the older literature generally reports botulinal toxin to be resistant to detoxification, our own work employing quantitative techniques, adequate numbers of experimental animals, and type A toxin of varying degrees of purity, has shown detoxification by trypsin and chymotrypsin to occur. Results with pepsin have been conflicting. The conclusion that must be drawn is that the capacity to act as an oral poison cannot be a matter of <u>absolute</u> resistance by the toxin to the activity of proteolytic enzymes in the intestinal environment (2). For lethality to be expressed, or any other action of the ingested toxin, it is only necessary for the smallest harmful amount or threshold effective dose of toxin to escape across the alimentary tract barriers before detoxification in the intestine has had time to proceed to completion.

In clinical cases of food poisoning, toxin is swallowed with a variety of different foods. It is conceivable that ingested foods can affect the oral potency of toxin by influencing the rate of intestinal detoxification, for example, by competition for or inhibition of proteolytic enzymes. We have been able to establish that the state of alimentation and the kinds of foods ingested with toxin do influence oral toxicity as measured by changes in LD_{50} lethal oral dose values (3). Foods and toxin were given to a mouse in separate <u>per os</u> injections. This procedure was adopted in preference to mixing food and toxin <u>in vitro</u> in order to insure that all results were the consequence of <u>in vivo</u> interactions exclusively. Foods may act to increase or decrease oral toxicity of a fixed quantity of ingested toxin. In TABLE 1, for example, it is demonstrated that olive oil and egg albumin can increase oral toxicity both in terms of increasing the rapidity of deaths and total number of individuals succumbing to a given quantity of toxin. This is unexpected if we have been thinking in terms of food competing for proteolytic enzymes in which case we would not predict an increase in toxicity in the presence of olive oil, a substance which does not react with proteolytic enzyme. Probably different foods can affect toxic potency by different mechanisms.

Incidentally, it is a part of the mythology of botulism that those who partake of alcoholic beverages at the fatal feast suffer less serious consequences than those who do not indulge. A few experiments, therefore, have been performed to test this belief. As can be seen in TABLE 2, ingestion of mixtures of brandy and egg albumin, a reasonable simulation of human experience, particularly at Xmas and New Year's parties, and mixtures of brandy with olive oil tended only to reverse slightly, if at all, the enhancing effect on toxicity of the food without reducing potency below the level experienced upon ingestion of toxin alone. Thus, the consumption of alcoholic beverages has no remarkable prophylactic value in botulism except insofar as the true imbiber eats less food. Taking into

account the weight difference between man and mouse at higher doses of brandy than those employed in the experiments performed, one is subject to the risk of drinking lethal quantities of brandy. In such an event, the anticipated cure might be more pleasant than the disease, but equally fatal.

Insufficient data are at hand to permit generalization and prediction of how particular kinds of foods will act. We do not know by what mechanisms the foods affect the toxic dose. But five possibilities worthy of investigation are self evident: foods might protect toxin against destructive intestinal influences; they might act to increase or decrease the secretion of digestive juices; they might combine with the toxin to form larger particles less able to penetrate the gut wall than is free toxin; they might modify the physiological bases of intestinal permeability; or they might have an effect on the rate of peristalsis with a consequent decrease or increase in the sojourn of the toxin in the part of the intestine offering the maximum opportunity for systemic absorption. No one of these possibilities has been adequately explored.

I suspect the effect of food is less by direct action on the toxin than on mechanisms influencing peristalsis and the permeability of the small intestine to whole protein. If foods did interfere with detoxifying proteolysis in the gut, one might hope to show an enhancement of oral toxicity by the use of specific inhibitors of enzyme, trypsin. Soybean and egg albumin trypsin inhibitors have been tried and both fail by their presence in the alimentary tract to affect the oral toxicity of the tetanus and crystalline botulinal toxins (TABLE 3) under the conditions of our tests. The total number of mice succumbing to varying doses of orally administered toxin was not found to be significantly different in the presence and absence of the trypsin inhibitor when a twofold dilution series of toxin was employed at dilutions somewhat above and below the oral LD_{50} dose. The inability of trypsin inhibitor to increase oral potency of the toxin was disappointing, since one would guess that a significant fraction of the great difference in the amount of toxin required for an oral lethal dose relative to a parenteral dose would be the result of destructive tryptic proteolysis in the gut which should be reversed by a specific enzyme inhibitor.

We will now turn our attention to the problems of permeability. How does the toxin go from the intestine to the blood stream? There is good evidence that the route taken is a lymphatic one (4,5), and that the lymphatic route is the only avenue (5). If the lymph draining the intestine is diverted from the body by cannulization of the thoracic duct so that none of the lymph can be spilled over into the blood stream, botulinal toxin fed an animal does not cause poisoning (5). This observation means two things, one, that absorption from the gut is limited to the lymphatic route and second, that the toxin poisons exclusively by direct action on extraintestinal tissue. These observations return a dividend of general scientific usefulness.

They provide us with an experimental means for deciding whether or not an oral toxin acts directly on the gut and legitimately can be considered a true enterotoxin. This question is not always easily answered. A case in point is the staphylococcal enterotoxin which traditionally has been thought to act on the intestinal tract directly, a point of view which has been placed in doubt by investigators at the University of Chicago. Since the staphylococcus toxin is a protein, it seems possible that a definitive answer should be forthcoming by observing animals orally fed the toxin and cannulated to prevent intestinal lymph from flowing into the general circulation.

Does the toxin that escapes from the alimentary tract and enters the blood stream actually have the dimensions of a protein? Such a question is related to the biochemical one of the size of the ultimate toxic particle.

Since in the natural situation botulism results from food poisoning, one might hope Nature to be parsimonious and to permit only the toxicologically active fragment of the protein particle to escape from the intestine into the general circulation. Such a possibility would be reinforced by any normal tendency of the alimentary tract's permeability barriers to refuse passage to whole protein. Heckly, Hildebrand and Lamanna (4) have studied this question. They have found the systemically absorbed toxin which appears first in the lymph and then in the blood to have the dimensions of a protein. By ultracentrifugal analysis, the sedimentation value (S_{20}) of the toxin appearing in the lymph draining from the small intestine of the rat was found to be 7.9 ± 3.5 which is within the size range of protein. Defeat has followed an attempt to find toxicity resident in a particle smaller in size than a protein.

The data available indicate that the toxin in lymph has dimensions within the range for proteins and need not be broken down to smaller non-protein elements in order to pass through the intestinal barriers into the lymphatic system. While the observed sedimentation coefficient of the absorbed toxin in lymph was significantly less than of crystalline toxin, it is probable that a small percentage of the crystalline toxin can dissociate to the smaller protein particle size, and it was this portion of toxic protein whose passage into the lymph was favored.

There is no evidence that crystalline toxin in the intestine is "digested" into smaller-sized toxic particles. This statement is based on sedimentation coefficients determined for crystalline toxin both before and after exposure to residence in the small intestine of the rat for a period of 2 hours. The particle size of the bulk of the toxin was not demonstrably reduced by exposure of the crystalline toxin to the digestive process in the living intestine. The sedimentation coefficient of the batch of crystalline toxin employed was in good agreement with the 17.3 value or 900,000 molecular weight reported for this material (6).

There is no reason to believe that the toxin crosses the small intestine as smaller than protein particles which are reaggregated in lymph to the dimensions of a protein. Rather our view is that the small intestine does not present an absolute barrier to the passage of protein. Botulinal toxin is but one among many whole proteins which can be absorbed from the small intestine in small quantities.

That the true particle size of toxin in lymph is not determinable by sedimentation studies because of absorption of the toxin to albumin is most unlikely. Since albumin is the most abundant and highly charged of the lymph proteins one might infer the toxin to be absorbed above pH 7 to albumin rather than to globulin. The fact that the toxin present in lymph migrates electrophoretically at the same rate as crystalline toxin, rather than at a rate corresponding to some value intermediate to toxin and albumin or to the value for albumin argues against the existence of a small molecular weight toxophore adsorbed to albumin. In addition, toxic lymph when dialyzed against serum albumin does not release toxic material able to pass across the walls of dialysis tubing. By placing mixtures of crystalline toxin and proteolytic enzymes in dialysis tubing one does not find toxic material escaping from the bag. This result might be expected to follow if proteolytic enzymes could chop off pieces of the protein molecule, and thus permit the escape of smaller-sized dialyzable toxophoric pieces. The conclusion to be drawn from these experiments (4) is that the measured sedimentation coefficients of toxin which has passed from the intestine into lymph are values for toxin unassociated with a carrier protein.

In another effort to settle the question of the size of the ultimate toxic particle, we have determined the sedimentation coefficient of type A botulinal toxin in lymph and blood after intravenous injection in rabbits (7). Such exposure of toxin to the *in vivo* extra-alimentary environment for as long as two hours did not reveal the occurrence of toxic materials with dimensions smaller than that of a protein. Thus residence in the body fluids did not demonstrably result in any disassociation or breakdown of the toxin to low molecular weight non-protein toxic moieties.

Alone among the classical bacterial exotoxins, botulinal toxin has been considered to be an oral poison. If botulinal toxin is truly unique in this respect, exploration of this situation might provide clues of a biochemical nature to the biological properties of the toxin. Unfortunately, oral toxicity is *not* a characteristic unique to botulinal toxin (8, 9). Both diphtheria and tetanus toxins, materials not associated with clinical cases of food poisoning, can act as oral poisons and at the very most are only one order of magnitude less toxic orally than botulinal toxin in terms of the number of intraperitoneal LD_{50} doses equivalent to one oral LD_{50} dose (TABLE 4). This finding suggests that oral toxicity of the bacterial exotoxins is not an expression of intrinsic qualities of chemical

structure of the toxins, but rather is a consequence of a physiological fact. This fact is the inability of the alimentary tract of the so-called normal animal to prevent the escape of small quantities of different kinds of whole proteins into the general circulation by way of the lymphatic route. The alimentary tract does not present an absolute barrier to the systemic absorption of whole protein, a fact which allergists have long recognized (10). The prime avenue of escape is the small intestine, probably the jejenum chiefly. Intrarectal instillation of toxin in monkeys (11) and rabbits (12) is slower than oral administration in causing deaths.

Potentially, any toxic protein is an oral poison if its potency is high enough for the minute amounts crossing the intestinal wall to exceed the threshold values for physiological activity at locations distant from the intestine. I emphasize the term minute quantity, since a lethal dose of botulinal toxin for the mouse involves fantastically small weights of material, the order of 1 to 10 thousandths of a millionth of a gram. The fact of escape of toxic protein through the wall of the large and small intestines should not shock us in spite of the classical teaching of physiology that the intestine is a formidable barrier to passage of protein. Hogben (13) has neatly stated a philosophy relevant to the problem of the penetration of tissue barriers by large-sized molecules such as microbial toxins: "Passage across cell membranes must be considered in statistical terms of likelihood and unlikelihood. Given a sufficiently sensitive method, any substance can be shown to cross a boundary". Even objects as large as non-pathogenic bacteria and yeast can pass from the intestine of normal rats to lymph though the numbers are extremely small (14). There is a possible correlation between the size of a particle and the number penetrating the gut wall since fewer yeasts escape than bacteria, and fewer of these large particles than toxin. Is such experience indicative of a similar path and mechanism of escape from the intestine for these qualitatively different kinds of particles, namely, diffusion from the intestine through "holes" in the intestine varying statistically in diameter in a normally distributed manner?

With bacterial toxins, for example the neurotoxins, extremely small rates of passage of proteins across tissue barriers can have pathological consequences. This means we cannot treat the passage of toxic proteins in the same vein as the physiologist, who, in considering permeability of tissues to proteins, is generally focusing his attention on orders of magnitude of penetration considerably beyond those of concern to the bacteriologist, immunologist and pathologist. Incidentally, the bacterial toxins can serve as useful tools to monitor the specificity of action of substances changing the permeability of the intestine to particular classes of compounds. For example, disodium ethylenediaminetetracetic acid (EDTA) increases absorption of heparin and heparinoids (15), highly charged anionic substances, but does so without increasing

passage of botulinal toxin in mice (16) or non-pathogenic bacteria in rats (14).

I should now like to turn my attention to the question of the oral dose of toxin required for poisoning the individual.

The opportunity for harm to befall the host, is a relative matter arising from the interaction of host factors and the harmful agent. In botulism the biochemical substrate of the peripheral nervous system of the host affected by the toxin may be in quantity independent of the body weight of the poisoned animal. This can follow from the fact that the number of nerve cells in an individual is fixed at birth and so does not increase with size and age of the individual. In mice lack of a relationship between body weight and the quantity of toxin required for a fatal parenteral dose has been found (17). The weight of toxin required for a fatal dose is the same for the small and large mouse. This is not a finding peculiar to botulism. We have found the same fact to be true for tetanus toxin (TABLE 5). Similar reports exist for Shigella paradysenteriae endotoxin in mice (18), β -naphthyl-thiourea in rats (19) and histamine in mice (20).

Of interest is the fact that the experience with a parenteral route of injection cannot be generalized to include the oral route. With type A crystalline botulinal toxin and tetanus toxin which we have tried, the youthful mouse required more toxin than did the older heavier mouse for a lethal dose (TABLES 6 and 7). While common sense might dictate a skeptical attitude toward such a finding, a fact of anatomy may justify the finding. In length the small intestine averages 40 cm in the young 12-14 g mouse, and 57 cm in the old 40-43 g mouse. There is, roughly speaking, 50 per cent more intestinal surface area provided for the systemic absorption of toxin in the large than the small mouse under comparison. Since the small and large mouse require the same minimum parenteral dose for lethality, the lesser oral dose for the large mouse could merely reflect the greater opportunity for systemic absorption before peristalsis removes ingested toxin from the bounds of the small intestine where absorption is most prominent.

No evidence has been developed for pinpointing the chemical molecular basis for toxicity in botulism, staphylococcal food poisoning and in infectious diarrheas. Oral toxicity for botulinal and other recognized food poisoning toxins can hardly be considered an unusual property because diphtheria and tetanus toxins, bacterial exotoxins not ordinarily thought of as oral poisons, will cause toxemia when ingested in sufficient quantity. Perhaps in diphtheria this fact has some role to play in the natural infection since the organisms growing in the naso-pharyngeal area are producing toxin which must in part be ingested as an inevitable consequence of the swallowing reflex. In some clinical cases of cryptic tetanus it

would be wise to seek for an unsuspected source of intestinal absorption of toxin.

Oral toxicity of any toxin would appear to be affected by any factor which can influence the length of residence of active toxin in the small intestine or the permeability of any portion of the gut to whole protein. Our understanding of these factors is still at the stage of development of fundamental descriptive data. We require detailed knowledge of the anatomical and physiological bases for toxicity by the oral route before we can achieve biochemical understanding of toxicity at the molecular level.

What general conclusions about bacterial exotoxins as oral poisons can we draw which has relevance to clinical medical experience? Unless a toxic protein produced by a bacterial species can directly adversely affect the normal activity of the alimentary tract, food poisoning should not be attributed to any unique intrinsic chemical properties which account for oral toxicity and are absent for other kinds of poisoning by bacterial toxins. The actual assessment of the capacity of a bacterial exotoxin to act as an oral poison must rest on an understanding of ecological circumstances: the relationship of food consumption to food preparation, and environmental factors influencing bacterial production of toxins in foods. These circumstances determine whether or not a particular organism will occur in a food and can grow to produce sufficient toxin to survive food preparation procedures such as cooking and to avoid total destruction in the intestine so that small quantities escaping the alimentary tract barriers by way of the lymphatic route are above the threshold values needed for pathological effects to manifest themselves. In this light it is sanitation, bacterial ecology, and the feeding habits of animals, and not biochemistry which have the stronger light to cast upon the calculation of the possibilities for the actual occurrence of clinical cases of food poisoning. This concept expands the list of harmful organisms which potentially can cause food poisoning. We can expect on rare occasions the proper interconnection of events which will result in cases of food poisoning by organisms not ordinarily believed to be food poisoning organisms.

The basis for oral toxicity would seem to rest on high potency of a toxin associated with a lack of capacity of the small intestine to prevent, in an absolute sense, the systemic absorption of proteins in small quantities. My basic hypothesis then, is that oral poisoning by a bacterial exotoxin is an accident of immediate circumstance. The historical biological origin of oral toxicity is not in an orderly evolution of proteins specifically directed toward the acquisition of unique properties conferring the character of oral toxicity.

LAMANNA

In reviewing what has been said, one must be impressed with the predominance of questions raised rather than solidly established facts offered. I have adopted this mode of presentation consciously. We are living in an age when it has become the habit of scientific institutions to sell themselves to the public. As a result the popular press is bombarding us with a continuous succession of scientific triumphs. This must sometimes have a discouraging effect upon the uninitiated scientist and students. By constantly praising ourselves, the rising generation may come to the feeling that it has grown up too late to participate in the real progress of science. The administrator may tighten his purse strings against the true need for research unless he can be promised a materialization of the breakthroughs he has been reading about. And so I chose my emphasis with an eye to satisfying the need to reassure our neighbors that we have not achieved true wisdom in all things. There remains many a thing both elementary and subtile to be learned even in such an old-fashioned subject field as is represented by the bacterial toxins.

In conclusion, I hope I have reassured you that if the food you eat poisons you, it is an accident and not a diabolical plot against you planned by Mother Nature. Good day and good eating to you.

REFERENCES

1. Schantz, E.J., D. Stefanye, and L. Spero, 1960. Observations on the fluorescences of botulinum toxin and other proteins denatured with urea. J. Biol. Chem. 235, 3489-3491.

2. Meyer, E.A. and C. Lamanna. 1959. Activity of Type A botulinal toxin and hemagglutinin exposed to proteolytic enzyme. J. Bacteriol., 78, 175-180.

3. Lamanna, C. and C.E. Meyers. 1960. Influence of ingested foods on the oral toxicity in mice of crystalline botulinal type A toxin. J. Bacteriol., 79, 406-410.

4. Heckly, R. J., G.J. Hildebrand, and C. Lamanna. 1960. On the size of the toxic particle passing the intestinal barrier in botulism. J. Exptl. Med., 111, 745-759.

5. May, A.J. and B.C. Whaler. 1958. The absorption of Clostridium botulinum type A toxin from the alimentary canal. Brit J. Exptl. Path., 39, 307-316.

6. Putnam, F. W., C. Lamanna, and D.C. Sharp. 1946. Molecular weight and homogeneity of crystalline botulinus A toxin. J. Biol. Chem., 165, 735-736.

7. Hildebrand, G. J., C. Lamanna, and R. J. Heckly. 1961. Distribution and particle size of type A botulinum toxin in body fluids of intravenously injected rabbits. Proc. Soc. Exptl. Biol. Med., 107, 284-289.

8. Lamanna, C. 1960. Toxicity of bacterial exotoxins by the oral route. Science 131, 1100-1101.

9. Lamanna, C. 1960. Oral poisoning by bacterial exotoxins exemplified in botulism. Ann. N.Y. Acad. Sc., 88, 1109-1114.

10. Ratner, B. and H. W. Gruehl. 1934. Passage of native protein through the normal gastrointestinal wall. J. Clin. Invest., 13, 517-532.

11. Dack, G. M. and D. Hoskins. 1942. Absorption of botulinum toxin from colon of Macaca mulatta. J. Infect. Dis., 71, 260-263.

12. Legronx, R., J.C. Levaditi, and C. Jeramec. 1945. Influence de voies d'introduction de la toxine sur le botulisme experimentale du lapin. Ann. inst. Pasteur., 71, 490-493.

13. Hogben, C.A.M. 1960. The alimentary tract. Ann Rev. Physiol. 22. 381-406.

14. Hildebrand, G. J. and C. Lamanna. 1961. Unpublished observations.

15. Cronheim, G. E. and E. Windsor. 1961. Gastro-intestinal absorption of heparin and synthetic heparinoids. Nature, 190, 263-264.

16. Lamanna, C. 1961. Unpublished observations.

17. Lamanna, C., W. I. Jensen, and I.D.J. Bross. 1955. Body weight as a factor in the response of mice to botulinal toxins. Am. J. Hyg., 62, 21-28.

18. Zahl, P. A., S. H. Hutner, and F.S. Cooper. 1943. Age as a factor in susceptibility of mice to the endotoxin of bacillary dysentery. Proc. Soc. Exptl. Biol. Med., 54, 137-139.

19. Rall, D. P. and W. C. North. 1953. Consideration of dose-weight relationships. Proc. Soc. Exptl. Biol. Med., 83,825-827.

20. Angelakos, E.T. 1960. Lack of a relationship between body weight and pharmacological effect exemplified by histamine toxicity in mice. Proc. Soc. Exptl. Biol. Med., 103, 296-298.

Table #1

Effect of per os injection of egg albumen & olive oil on the toxicity of botulinal toxin

Cumulative Deaths after Receiving Toxin per Os									
Toxin Dilution	Hr	Sham	Albumen	Olive Oil	Toxin Dilution	Hr	Sham	Albumen	Olive Oil
1:2	12	1	6	9	1:8	12	2	4	3
	24	7	16	15		24	6	9	8
	36	10	20	18		36	7	14	12
	48	12	20	20		48	9	15	13
	48	17	20	20		48	11	20	16
1:4	12	0	5	3	1:16	12	2	1	0
	24	1	12	6		24	3	1	4
	36	2	14	9		36	3	2	6
	48	4	16	10		48	4	2	6
	48	9	18	14		48	5	12	6
Total Deaths..................................							42	70	56

(From 3)

Table #2

Effect of brandy (96 proof) on the per os toxicity of botulinal toxin in the presence of food

(a) Presence of olive oil*

Cumulative Deaths after Ingestion of Toxin									
Hr	Undiluted			1:2			1:4		
	Control	Olive Oil	Olive Oil & Brandy	Control	Olive Oil	Olive Oil & Brandy	Control	Olive Oil	Olive Oil & Brandy
6	2	4	3	2	0	3	1	4	1
12	5	15	8	3	1	4	3	4	2
18	12	17	12	4	4	5	5	4	2
24	15	19	12	9	9	9	6	6	3
30	18	20	16	12	12	10	8	10	7
36	19		19	15	12	11	9	12	8
42	19		20	16	12	12	9	14	10
48	20			16	13	13	10	15	13
60				17	15	14	11	15	14
72				18	15	14	11	15	15
72					16	15	13	17	17
Total Deaths..................................							51	53	52

(b) Presence of egg albumen

Continued on next page

TABLE #2 Continued

Hr	Cumulative Deaths at Hr Indicated after Toxin Ingestion								
	Dilution of Toxin								
	1:1.5			1:2			1:4		
	Sham	Albumen	Albumen & Brandy	Sham	Albumen	Albumen & Brandy	Sham	Albumen	Albumen &Brandy
12	1	3	1	1	2	0	1	1	0
24	2	6	6	2	2	2	1	2	2
36	2	12	8	3	5	2	2	3	3
48	4	14	10	3	5	3	2	4	4
60	5	15	10	3	7	4	2	4	4
72	5	15	10	4	7	4	2	4	4
72	5	15	13	4	9	5	2	4	4
Total deaths......................................							11	28	22

(From 3)

Table #3

NUMBER OF MICE SUCCUMBING TO ORALLY ADMINISTERED CRYSTALLINE TYPE A BOTULINAL TOXIN IN THE PRESENCE AND ABSENCE OF ORALLY ADMINISTERED TRYPSIN INHIBITOR

Experiment	1	1	2	3	1	2	1
	Soybean trypsin inhibitor						
Inhibitor present per mouse	0.2mg.11/32*	0.5mg. 3/12	2/12	9/20	1mg23/32	9/32	1.5mg12/32
Inhibitor absent	12/32	3/12	4/12	11/20	25/32	8/32	13/32
	Eggwhite trypsin inhibitor						
Inhibitor present per mouse	0.5mg.8/32	1mg. 22/32	10/32	1/32	2mg6/32		
Inhibitor absent	11/32	21/32	8/32	4/32	8/32		

*Dead mice/number tested.

(From 9)

TABLE #4

NUMBER OF INTRAPERITONEAL LD_{50} REQUIRED FOR ONE ORAL LD_{50}*

Species		Toxin
Mouse, 20 gm.	Botulinal type A	50,000 to 250,000
Mouse, 20 gm.	Tetanus	80,000 to 1,200,000
Guinea Pig 900 gm.	Diphtheria	100,000

*With the mice the toxins were administered orally by the use of a slightly curved blunt-nosed needle on a syringe. Diphtheria toxin was administered by forced feeding of gelatin capsules containing concentrated toxin. These methods appear to successfully introduce toxin into the gut without contamination of the mouth and throat and appear to be well tolerated without evidence of tissue trauma. Deaths, therefore, are thought to be truly representative of absorption of toxin from the normal gut.

(From 9)

TABLE #5

INTRAPERITONEAL LD_{50} OF TETANUS TOXIN SOLUTION WITH DIFFERENT WEIGHT MICE*

Average Weight: Small Mice (gm)	Average Weight: Large Mice (gm)	Technician	LD_{50}: Small mice	LD_{50}: Large mice
7.6	39.3	A	1,631,000	1,350,000
9	40	A	189,000	244,000
9.7	40.1	A	364,000	283,000
9.4	39.6	A	389,000	305,000
11	39	A	111,000	55,000
		B	81,000	53,000
9.4	37	A	257,000	259,000
		B	323,000	212,000

* The titration values for the small and large mice are not significantly different, the variation being within the limits of experimental error. The method of Pizzi (1950), which permits the use of the Reed and Muench type of calculation, was employed for determining the standard error of the LD_{50}.

(From 9)

TABLE #6

TITRATIONS BY THE ORAL ROUTE OF CRYSTALLINE BOTULINAL TYPE A TOXIN IN MICE OF DIFFERENT WEIGHTS

Toxin (ml.)	Weight of mice (grams)									
	Experiment 1		Exper. 2		Exper. 3			Exper. 4		
	13-15	38-40	12-14	40-43	13-14	20-22	37-39	13-14	20-22	34-36
0.08					20/20	17/20		12/20	15/20	
0.04	19/20*	20/20			17/20	13/20	17/20	5/20	9/20	13/20
0.02	2/20	20/20	7/20	16/20	2/20	8/20	11/20	1/20	1/20	11/20
0.01	1/20	9/20	1/20	15/20	8/20	5/20	8/20	0/20	2/20	10/20
0.005	2/20	5/20	3/20	7/20	1/20	3/20	7/20	1/20	1/20	0/20
0.0025	1/20	3/20	1/20	3/20			2/20			0/20
0.00125			0/20	2/20						

* Dead mice/number injected. The difference in deaths between the 13-14 gm. and 20-22 gm. mice is probably not significant. On the other hand, the LD_{50} for the largest mice is significantly less (less toxin required for death) than for the smaller mice.

(From 9)

TABLE #7
ORAL TOXICITY OF TETANUS TOXIN FOR MALE MICE OF DIFFERENT WEIGHTS

Toxin (ml.)	Weight of mice (grams)					
	Experiment 1		Experiment 2		Experiment 3	
	13-14	37-39	13-14	38-40	13-14	34-35
0.75			2/12	10/12	4/12	8/12
0.50	5/12*	6/12	2/12	4/12	5/12	3/12
0.25	2/12	1/12	0/12	3/12	1/12	4/12
0.125	0/12	1/12	0/12	4/12	0/12	2/12
0.0625	0/12	2/12				
0.03125	0/12	0/12				

*Dead mice/number injected. (From 9)

STATIC FATIGUE FRACTURING OF HIGH PRESSURE GAS VESSELS

F. R. LARSON
WATERTOWN ARSENAL
WATERTOWN, MASS.

INTRODUCTION

Delayed fracturing of service components has always been a great problem. This is primarily because of the unpredictable nature of the failures. Usually these failures occur at stresses well below the yield strength of the material used. In most cases of delayed failures, the fracturing event is quite catastrophic and the brittle nature of the material used is evident. It is therefore clear that the basic toughness of the metal employed plays a major role. However, not all delayed fractures are caused by low toughness materials; only that these materials are more prone to this type of failure. This is because small cracks can have pronounced effects on the behavior of low toughness materials. The phenomena is best illustrated by figure one which is a schematic derived from the Griffith[1] fracture theory. Briefly, this figure shows the "fracture boundary" for two materials and illustrates that a material containing a crack can be only loaded to a certain stress, beyond which unstable rapid fracture occurs. The main point to notice is the limited crack-size load-area for the low toughness material as compared to the high toughness material.

Assuming a structure has been fabricated, inspected, proof tested, and put into service, it would be safe to state that this structure would be within the fracture boundary and should give satisfactory service. In the majority of cases this is so but occasionally service failures do occur. It then is clear that one of three things could have happened. The first is the slow growth of cracks or an increase in crack size. The second is a decrease in toughness of the material employed and although the toughness of a material is an inherent property dependent upon prior processing history and does not usually change with age, it is markedly sensitive to service temperature. Finally, the part could have failed due to an overload. The third is in the realm of design and service

limits but could occur as a result of an accidental overload or poor design; however, it is not the purpose of this presentation to discuss this phase of the problem. Furthermore, it is not intended to discuss fracture mechanics in detail for it is a complex subject. The main point which is of concern is the crack size toughness relationship as illustrated.

Assuming that the basic toughness of a material does not change with service, except for temperature effects which will be discussed later, it is clear that cracks must grow with service. Structural failures often occur when there are cyclic stresses. The fatigue of parts involved, usually in the region of discontinuities, relates to stress concentrations which cause progressive cracking. For example, failures in liberty type ships at sea have occurred and considerable information has been accumulated on these failures.[2] Many other similar type failures have occurred wherein it has been possible to connect the fatigue crack propagation with the cyclic stress. In fact in many cases it is possible to pin point the number of cycles of stress by the fracture surface markings. The important point is that a stress cycling causes propagation of the fatigue crack which grows to the critical size, and catastrophic fracture ensues. The major features of this type of failure are some sort of a stress concentrator or a discontinuity and a cyclic stress. The crack usually initiates at this stress concentrator and propagates a finite amount for each stress cycle. From examination of these fractures and even during the service of such components the cracks become evident. Therefore, it is possible through periodic examination by nondestructive testing techniques, such as magnaflux, to detect progressive damage of the part. The main point being that the damage is evident and can sometimes be detected before final failure.

The second important type of failure in this class is stress corrosion cracking[3]. In this case there is a corrosion environment which propagates the crack. It is not necessary to have cyclic stresses under these conditions. Stress corrosion cracking can be induced under certain conditions in most metals. One of the elements present in this type of failure is the growth of a crack caused by corrosive environment usually along a shear stress trajectory. It is, therefore, evident that any corrosive action, in the vicinity of high stresses, on the service components can be serious. By removal of the corrosive environment the stress corrosion cracking problem can be eliminated. It is then possible from examination of the surfaces involved in the stress corrosion phenomena to prevent or guard against this type of failure. Characteristics of the fracture surface in stress corrosion cracking failures usually are associated with an intergranular attack. The main feature of the crack, examined in cross section, is the many branches in areas adjacent to the main crack. Occasionally the fracture surface is very granular in nature and has a bright appearance.

A third class of failures is usually associated with hydrogen and this phenomena is commonly called hydrogen embrittlement[4]. The main characteristic of hydrogen embrittlement is that hydrogen migrates to an area of high stress and causes a low stress crack growth. The basic mechanism of hydrogen embrittlement has not yet been clearly defined and there is considerable argument about the actual detail of the atomic mechanism. Hydrogen may be present in the metal to begin with because of prior melting and thermal mechanical history. The presence of hydrogen usually causes a severe embrittlement particularly in high strength material. It is known that hydrogen can cause delayed fracturing at stresses that are as low as 30 percent of the yield strength and will usually occur in the order of 10 to 100 hours.

There is a fourth possibility which will be the main theme of this paper. It consists of a combination of the second and third category of fractures. Basically what happens is a small fairly inactive corrosion occurs and the by-product of this corrosive action is hydrogen. The hydrogen is charged into metallic parts surrounding the region of the corrosive action. The slow charging of hydrogen produces an embrittlement which is an extremely long time effect. The main damage is done by the charging of hydrogen; the by-product of the corrosion. This gas builds up with time, depending on the corrosion rate, and once a critical amount of hydrogen is reached, the low stress slow crack grows into the metallic matrix. Once the crack has reached the critical size, catastrophic failure ensues.

From the foregoing remarks it can be recognized that the failures of certain types of service components are a complicated and many prong problem. To analyze any service failure, the complete knowledge of events prior to and existing at the time of failure, is needed. It is the purpose of this paper to discuss a recent failure which is believed to be a classical example of static fatigue fracturing. It occurred in a pressure vessel at a stress which was approximately 30 percent of the yield strength of the material. There were no cyclic stresses involved, corrosion was evident, low toughness in the region of service temperature was determined from laboratory tests, so that all the elements of the delayed fracture in service were present prior to placing these components in the field.

BACKGROUND

On or about 9 September 1960, a pressure vessel in a missile that was in storage at Red River Arsenal exploded, completely destroying the missile.[5,6,7] The pressure vessel in question is part of a hydraulic accumulator system used to actuate the hydraulic servosystem. This vessel contains nitrogen gas at a pressure of approximately 4,000 pounds per square inch. The flange at one end of this vessel is screwed into a component called an oil dome. Between the pressure vessel and the oil dome is a cavity which contains a rubber bladder and hydraulic oil.

In operation the 4,000 psi nitrogen gas is released against the bladder which in turn pressurizes the hydraulic oil to 2,000 psi. This pressurized oil then drives the guidance system. The important fact is that the pressure vessel is charged to 4,000 psi with nitrogen gas at the manufacturer's plant and remains at that pressure until the missile is fired. This is an important step forward because previous missile systems have required pressurizing of the gas cylinders prior to firing the missile, so that the missile launchings were delayed for a period of 15 or 20 minutes. In this particular missile system it is possible to have the missile ready to fire in a couple of minutes. The tactical advantage is quite obvious.

RESULTS OF INVESTIGATION

To understand more clearly the nature of the failure, it would be helpful to review some details of the processing of the pressure vessels and the material used. The pressure vessel is made up of two parts, one is a forging and the other is hydroformed from plate stock. The two parts are called the flange end and the dome end. These parts are made from material which is an air hardening-die steel purchased to an Aeronautical Material Specification AMS 6437, popularly known as H-11. The flange and dome end are rough-machined on the exterior dimension and finish-machined internally and are joined by making a circumferential J-weld, using argon as a shielding gas with H-11 as a filler wire.

Subsequent to welding, bright hardening is done by austenitizing at 1850°F in a reducing atmosphere. The vessel is then air cooled, followed by triple temper at 1080, 1100, and 1120°F in order to obtain a Rockwell "C" hardness of approximately 47. This hardness corresponds to a 200,000 to 220,000 yield strength material. Following this heat treatment the welded vessel is finish-machined and proof pressure tested to approximately 8,000 pounds per square inch using hydraulic oil. The oil from the proof testing is then cleaned out of the vessel with alcohol. Prior to placing the vessel in service, it is charged with nitrogen gas which has a -80°F dew point. There has been some question regarding the quality control of this nitrogen gas as several of the vessels have been shown to have had considerable moisture on the interior surface. The vessel has, of course, one atmosphere of air in it because it was not evacuated prior to charging with nitrogen; thus, whatever moisture was contained in the air was also present internally after charging. The pressurizing of the nitrogen and air undoubtedly condensed some of the moisture present.

The first failure occurred at Red River Arsenal on or prior to 9 September 1960. The missile was in storage in an iglo and the pressure vessel ruptured catastrophically by shattering into many pieces, completely destroying the missile. The manufacturer's record indicates this vessel was charged on 7 April 1960. This would make

the total elapsed time between charging and failure about 22 weeks. Figure 2 is a photograph of the fractured segments from the failure. Chronologically the events are quite similar for a second failure which occurred on 7 October 1960 at Fort Bliss, Texas. Fragments from both vessels were shipped to Watertown Arsenal Laboratories and an investigation was undertaken to try to determine the cause of failure.

Both of the fractures were very similar so they will be discussed together, noting the minor differences. The actual location of fracture origin was made by an examination of the fracture surface markings commonly called chevrons which point back to a darker area with a little indentation on the inner surface. Figure 3 illustrates this for one of the vessels. Both fracture origins were about 3/8" from the weld. Visual examination revealed that neither origin was associated with the weld or the heat affected zone. It is probable that these fracture origins were in this location because the hoop stress is maximum in this region. The main characteristic of the fracture was a pit on the internal surface which was undoubtedly associated with a corrosion process. Surrounding the pit was a dark black area of slow crack propagation followed by an area marked with chevrons which defines the rapid fracturing stage.

An examination of the many fragments showed that the fracture was quite flat and the shear lips along the edges adjacent to the free surfaces were very small. These characteristics indicate the material was quite brittle. Because of the pit areas in the fracture nucleus, a more detailed study of the internal surface was made. The importance of determining the nature of the fracture origin is self-evident as previously pointed out. In the interior of the two fractured vessels, considerable amounts of corrosion products were present. Several areas had large flaky build-ups which were evidently corrosion products. These build-ups were probed with a needle and revealed no solid particles. X-ray analysis of these particles indicated they were mainly iron oxide. Some chromium oxide was also detected. In several areas underneath these build-ups were deep craters or pits.

The tendency for corrosion varies with material and within a given material with microstructure and strength level. The extremely low corrosion resistance of steel when heat treated above 200,000 psi is well known. This was further substantiated and evident from metallographic specimens which were left overnight. In the preparation of the specimens for the microstructure studies, some specimens were mounted, surface ground and then left overnight in air. In the morning one of the specimens had several deep pits with dark grayish, blue color around it. In order to determine whether or not stress corrosion cracking was one of the mechanisms involved, fragments from both failures were subjected to Magnaglow to determine if any branching cracks were present. The results were negative,

indicating that no cracks other than the main flat fracture were present. The microscopic examination of several areas which had several deep pits in the region of maximum hoop stress also revealed no cracks. It is safe to say that, stress corrosion cracking mechanism did not play a role in this failure.

To determine the mechanical properties of the failed components both ASTM standard diameter tensile and sub-size impact specimens were machined from the fractured parts. Along with the two failed service vessels, a third vessel which had been shown to have a lower hardness was also studied. Tables I and II present this data. The tensile properties, Table I, are normal for the heat treatment employed. From Table II it can be seen that this material is quite brittle having impact energy of about 2 ft-lb with 5 to 10 percent fibrosity. These low values of fibrosity indicate the brittle to ductile transition to be somewhere above room temperature; a very serious condition indicating a low toughness material. Complete chemical analysis was conducted on the failed parts. All elements were within their chemical range and no deviations were noted.

The microstructures were essentially tempered martensite with the softer vessel showing a much higher degree of tempering. During the metallographic examination of specimens for general microstructure, it was observed that there was a general tendency for grain boundaries to etch dark. Since it is known that this dark etching tendency is usually associated with a grain boundary embrittlement, a special reagent was used to further reveal this condition. With this reagent the delineation of grain boundaries was evident in all specimens that were examined as shown by Figure 4. It is well known that grain boundary embrittlement is usually caused by precipitation of carbides in the form of a film along the grain boundaries and can give rise to high transition temperatures and severe embrittlement. Further information about the probable cause of this grain boundary condition can be had from the isothermal transformation diagram for this steel as shown in Figure 5. Although this is an air hardening steel, it is necessary to cool it fairly rapidly to avoid the carbide nose as illustrated by the dotted line. Different cooling rates were applied to some specimens in an experiment to verify this. It was proven that the grain boundary condition results from slow cooling and is probably a carbide precipitate. The microstructures of several vessels taken from production clearly show these characteristics of carbide grain boundary precipitates. Further microstructural studies of the path of fracture, Figure 6, showed that the early stages of fracture were primarily intercrystalline. There was also evidence of intercrystalline cracks of about one half a grain diameter propagating out from the main fracture path. This intercrystalline cracking is similar to the phenomenon of temper embrittleness and is an extremely dangerous condition. All of the characteristics of high transition temperature (above room temperature) and low fracture toughness are evident in these tests. From

this information it is evident that the pressure vessel material had a fracture toughness transition above room temperature. If you recall, the pressure vessels were charged with gas, in April, in the summer time when ambient temperatures were high. The actual service failures occurred during the beginning of the cold season and only a very slight shift in ambient temperatures surrounding the components would be sufficient to cause a marked drop in the basic toughness of the material.

DISCUSSION

It is now clear that several events were contributory to this static fatigue fracture. First, and probably one of the major factors was the extremely brittle nature of the steel used; and the fact that this steel had a transition temperature above the service temperature. Second, the vessel was placed into service combining small amounts of moisture and possibly some corrosive media because of a plating process which is used. During the storage life, under pressure, a corrosive action was developed. Hydrogen was evolved from this corrosion reaction and charged into the base metal. The charging of the hydrogen into the base metal caused a slow crack growth. The dark surface in the region of the slow growth stage of cracking is quite common for those cases of hydrogen embrittlement. While the surrounding ambient temperature was high, the material could withstand fairly large cracks. However, as time moved into the colder months of the year, the lower ambient temperature caused a lowering of fracture toughness of the material. Since each vessel passed a proof testing of twice the service pressure, it is obvious that some sort of progressive deterioration had occurred. Referring back to Figure 1 it can be readily seen that both a slow crack was growing and because of a lower temperature the "fracture boundary" was moving to the left. Therefore, brittle fracture was inevitable.

The important point is that the material used was extremely brittle; and hence, the susceptibility to fracture by a shattering mode of failure in the presence of a crack or pit was very high. Other factors such as the slow growth of cracks and defects were contributory. Furthermore, the slow growth of these defects is known to be more severe in a material with low fracture toughness. Since the corrosion occurred on the internal surface of this component, it was not possible to employ inspection techniques which could detect progressive corrosion damage. Even if such an inspection procedure could be employed, it is highly unlikely that it would be possible to detect these cracks and flaws in any practical way because of their small size. The main conclusion is that the only safe alternative method is to fabricate these components from a material which has a high degree of toughness. To prevent failures of this type, it is possible to use certain basic guide lines and work in this direction to eliminate all the possible contributory factors.

The first and main important fact is, of course, the elimination of the brittle material and replacement with a material of higher degree of toughness at low temperatures as previously stated. To correct the situation that occurred in this failure, the following recommendations were made:

1. Depressurize all vessels in the field.

2. Replace all vessels fabricated with H-11 with new vessels made with AISI 4340 steel. This recommendation was necessary because a reheat treatment was not feasible since tests show that even at a much lower yield strength the H-11 material does not have adequate toughness.

3. To obtain the best combination of properties in the 4340 material, it is desirable to utilize a 160,000 to 180,000 yield strength range rather than the 200,000 yield strength range. With the introduction of this new material and strength level it will still be necessary to provide the best possible fabrication inspection techniques to insure a sound, crack free product.

4. Eliminate all processing that tends to charge hydrogen into these vessels during fabrication because it is not just the actual individual amounts of hydrogen that are present but the sum total of all hydrogen present that causes hydrogen embrittlement failure. These vessels should be evacuated and flushed with dry gas and all efforts should be made to prevent any moisture or corrosion causing medium in the internal cavity.

The above recommendations were adopted and all of the H-11 vessels were replaced with vessels made of AISI 4340 manufactured according to recommendations. As part of the quality control program several of these vessels have been periodically burst tested and an inspection of the internal surface made. There has been no evidence of corrosion and all burst tested vessels have behaved in a ductile manner even at burst temperatures of -65°F. Figure 7 illustrates this behavior. Notice the large distortion and shear lips, evidence of ductility and toughness as compared to the brittle fracturing of a H-11 vessel tested under similar conditions, Figure 8.

These new AISI 4340 vessels have been in service for over a year and in climatic conditions of severe cold and there have been no failures reported. Therefore, it is clear that static fatigue fracturing can be prevented by the intelligent use of materials and the utilization of good manufacturing control.

REFERENCES

1. GRIFFITH, A. A. "Theory of Rupture" First International Congress, Applied Mechanics Deft (1924) p 55.

2. JOHANSON, F. "Brittle Fracture and Fatigue in Ships" "Fatigue and Fracture of Metals" Book, John Wiley, 1952.

3. "Symposium on Stress Corrosion Cracking of Metals" Joint ASTM-AIME Meeting, Special Publication (1944).

4. TROIANO, A. R. "The Role of Hydrogen and Other Interstitials in the Mechanical Behavior of Metals" Trans ASM, v. 52 (1960) p 54.

5. LARSON, F. R. "Metallurgical Examination of Hydraulic Accumulator Failure" WAL TN 781.1/1, Mar 1961.

6. LARSON, F. R. "Metallurgical Evaluation of AISI 4340 Steel for Hydraulic Accumulator Service" WAL TN 781.1/2, Mar 1961.

7. LARSON F. R. and WILLIAMS, E. E. "Metallurgical Evaluation of AMS 6415E (AISI 4340) Forgings and Billets for use in the Hawk Hydraulic Accumulator" WAL TN 781.1/4.

TABLE I

TENSILE TEST RESULTS

No. of Hydraulic Accumulator	*Mark	Diameter (inch)	Y. S. 0.1% (psi)	Y. S. 0.2% (psi)	U.T.S. (psi)	Elong. (%)	R.A. (%)
Y295	F-T1	0.113	207,500	221,000	243,000	8.9	33.5
	-T2	0.113	203,500	217,500	251,000	6.7	32.1
	-L1	0.113	201,500	215,500	247,500	6.7	35.0
	-L2	0.113	207,500	222,000	253,000	8.9	35.0
	H-T1	0.111	248,400	254,100	258,300	4.4	32.3
	-T2	0.100	203,800	223,000	255,100	4.4	25.6
	-L1	0.113	204,000	222,000	255,000	6.7	32.1
	-L2	0.110	206,800	223,200	255,800	6.7	31.6
Y186	F-L	0.113	137,000	148,000	175,000	11.1	51.9
	-T	0.077	159,600	164,900	176,600	8.9	53.2
	H-L	0.113	143,000	150,000	175,000	13.3	47.2
	-T	0.109	161,300	163,400	177,400	13.3	51.6
Y903	F-T	0.113	207,000	231,000	260,000	6.7	27.6
	-L	0.138	214,000	229,000	265,300	6.7	28.0
	H-T	0.113	233,000	244,000	263,000	6.7	32.1
	-L	0.113	232,500	245,000	271,000	6.7	32.1

*F = Forged
H = Hydroformed
L = Longitudinal
T = Transverse

TABLE II

*CHARPY V NOTCH IMPACT RESULTS

No. of Hydraulic Accumulator	**Mark	Test Temp. (°C)	Energy (ft-lb)	Fracture % Fibrous
Y295	F-L1	20	2.0	5
	-T1	20	2.0	5
	H-L1	20	2.2	10
	-T1	20	1.5	5
	F-L2	-40	1.8	5
	-T2	-40	2.0	5
	H-L2	-40	2.0	5
	-T2	-40	1.5	5
Y186	F-L	24	4.7	65
	-T	24	4.3	55
	H-L	24	4.2	60
	-T	24	4.0	60
Y903	F-L	24	2.0	5
	-T	24	1.4	5
	H-L	24	2.0	5
	-T	24	2.2	5

*Type specimen employed:
0.394 x 0.118 x 1.125" with a notch depth of 0.024" and notch radius of 0.010".

**F = Forged
H = Hydroformed
L = Longitudinal
T = Transverse

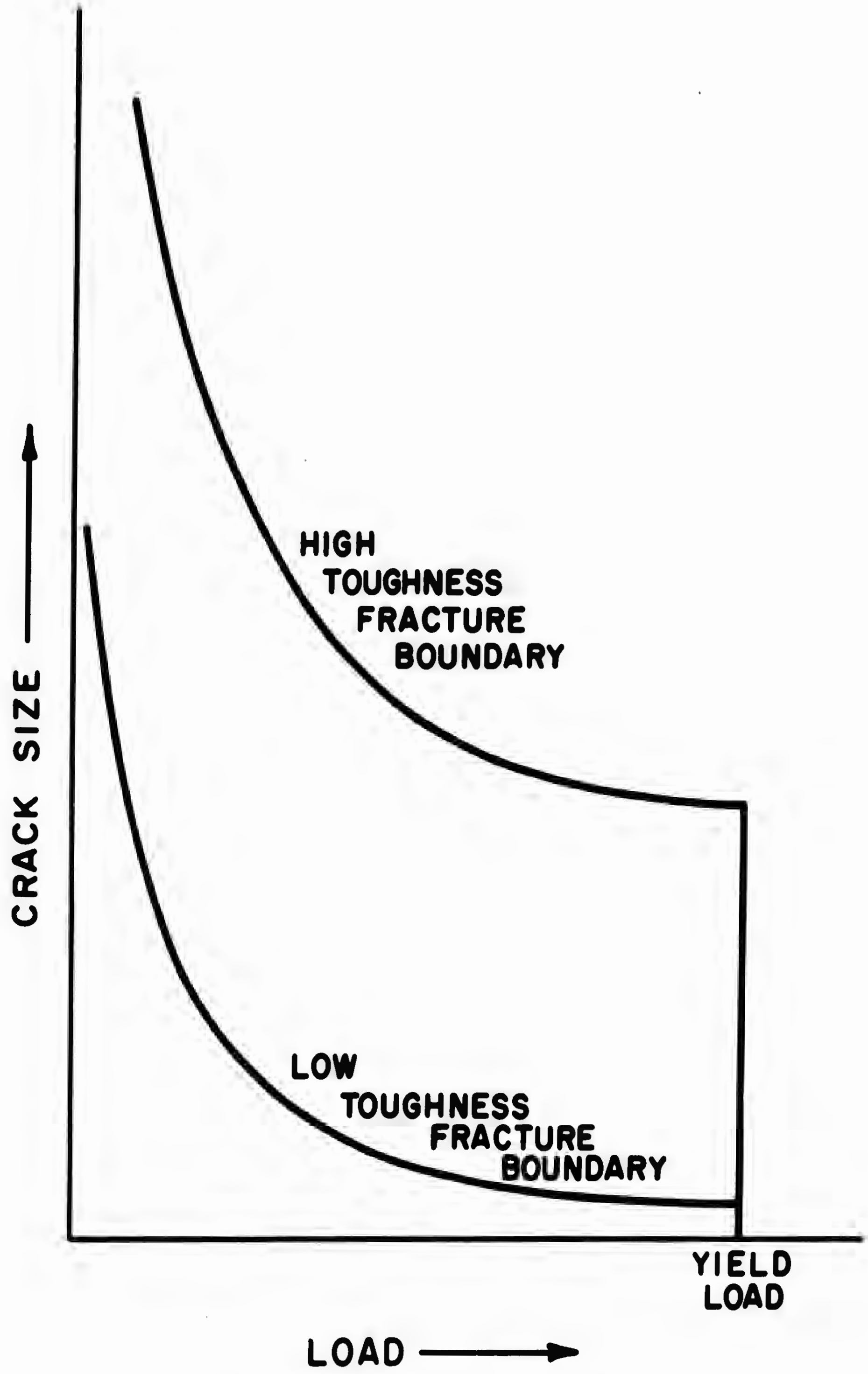


FIGURE I. STRUCTURAL LOAD-CRACK SIZE FRACTURE BOUNDARY FOR TWO MATERIALS.

19-066-257/ORD-62


INTEGRAL FLANGE
ELECTRIC SQUIB
COMPRESSED GAS CYLINDER
HYDRAULIC ACCUMULATOR FAILURE
19-066-1681/ORD-60

FIGURE 2

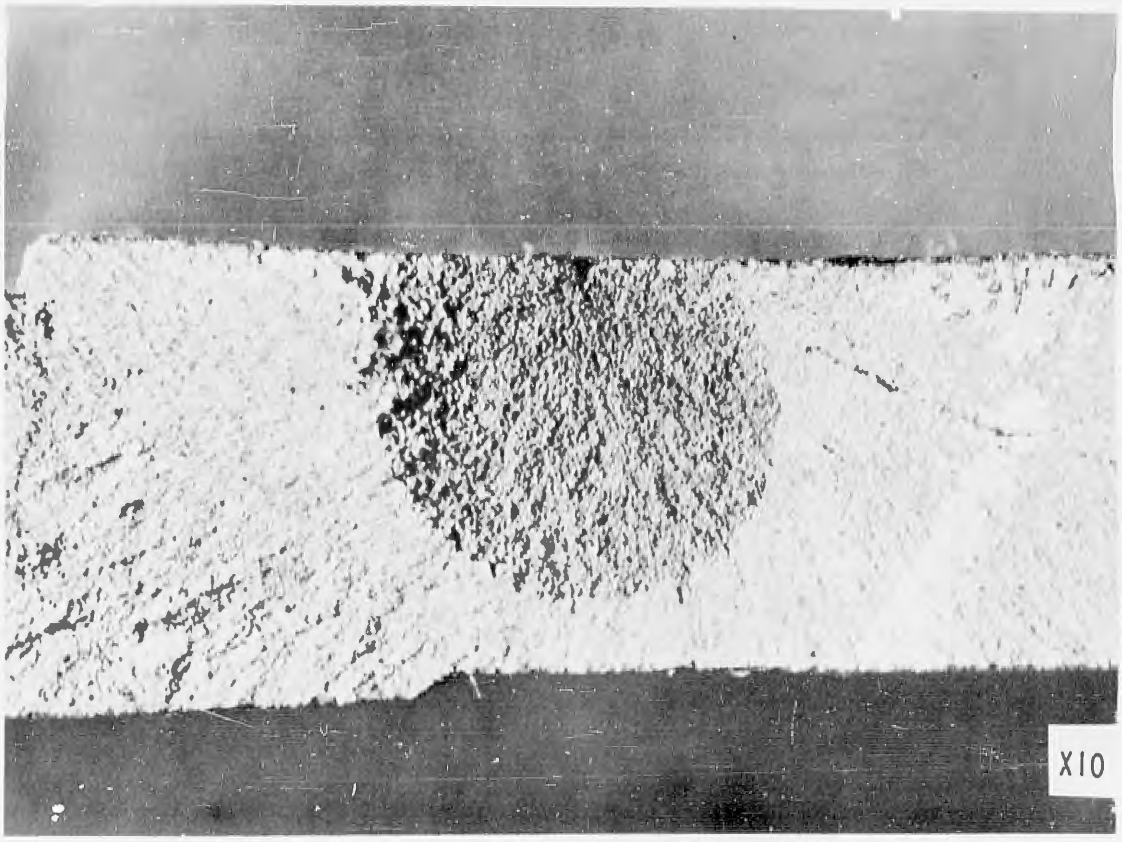


FIGURE 3: ORIGIN OF FRACTURE IN VESSEL (y295)

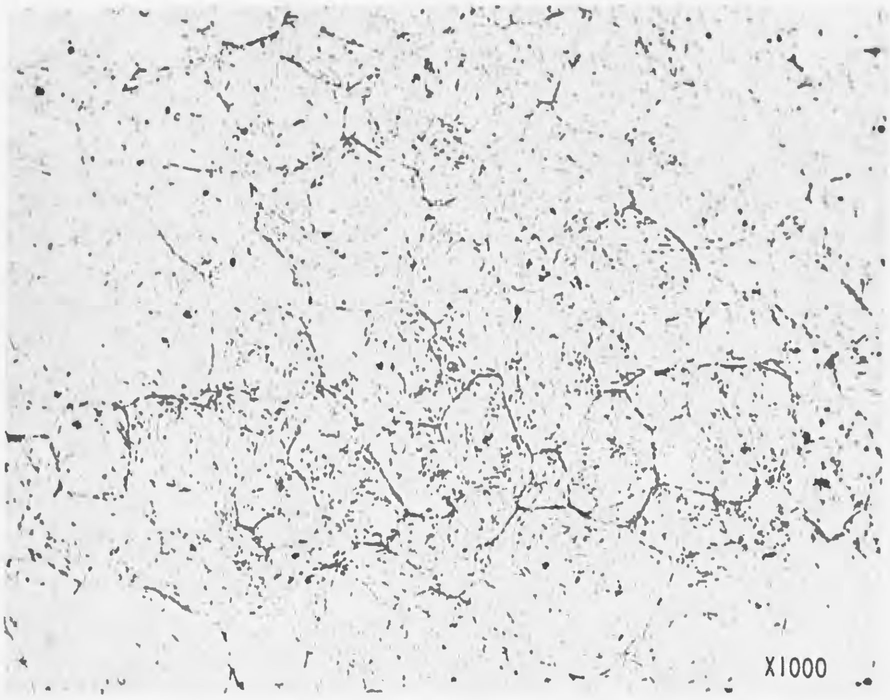


FIGURE 4: MICROPHOTOGRAPH OF GRAIN BOUNDARY CONDITION IN HYDROFORMED SECTION (y295) (TEMPER EMBRITTLEMENT ETCH)

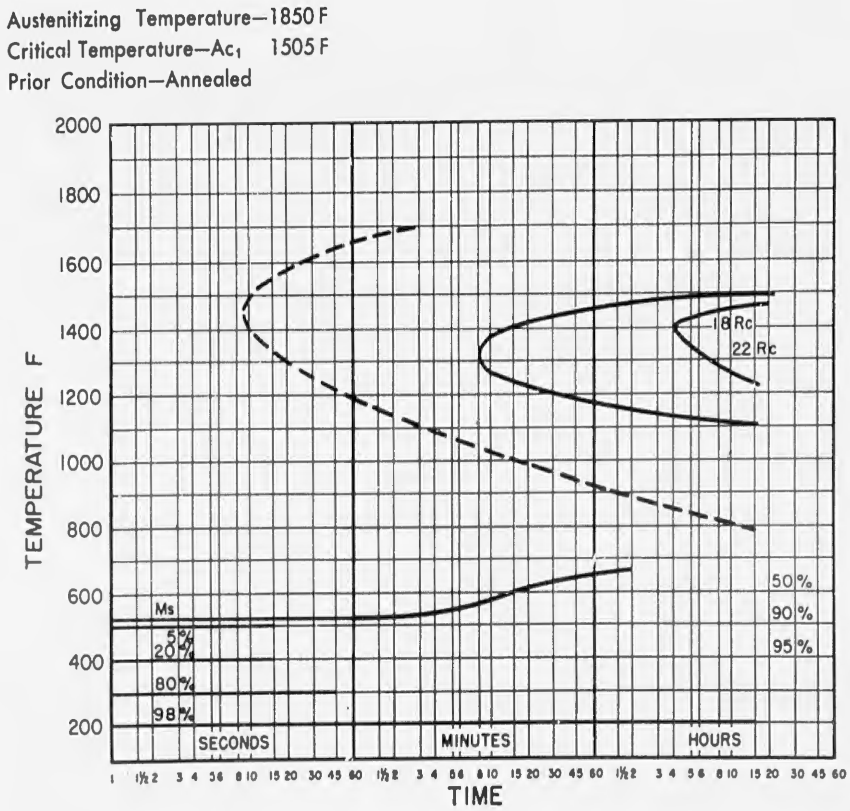


FIGURE 5: ISOTHERMAL TRANSFORMATION DIAGRAM FOR H-11 STEEL
(NOTICE DOTTED LINE INDICATES CARBIDE PRECIPITATION ZONE)

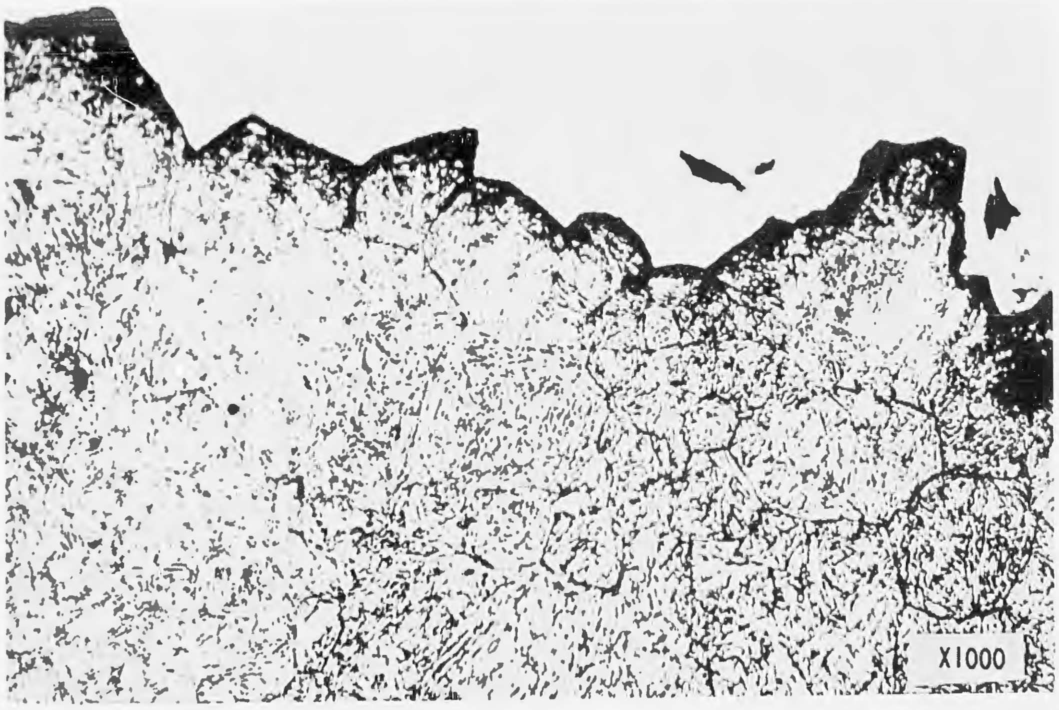


FIGURE 6: MICROPHOTOGRAPH OF FRACTURE PATH NEAR ORIGIN
(y295) - MOSTLY INTERGRANULAR

FIGURE 7: 4340 BURST-TESTED VESSEL

FIGURE 8: H-11 BURST-TESTED VESSEL

MATERIALS FOR BIOMECHANICAL APPLICATIONS

FRED LEONARD, CARL A. NIELSON, and JOSHUA NELSON
U.S. ARMY PROSTHETICS RESEARCH LABORATORY
WALTER REED ARMY MEDICAL CENTER
WASHINGTON 12, D. C.

INTRODUCTION

Recent advances in surgery have enhanced the need for providing polymeric materials which can be utilized in replacing or repairing tissues or organs which have been damaged as a result of trauma or disease. These materials have been designated materials for biomechanical applications.

The wide range of mechanical properties available in polymers makes them attractive candidates for such uses. However, even though a polymer may possess the desired physical properties, there is no assurance that it may be successfully used in the body. Biological receptivity is the sine qua non for the long term utilization of a surgical repair material.

Why some polymeric materials are biologically receptive and others are not is not known with certainty at present.

Some investigators favor the view that tissue receptivity is related to polymeric molecular structure (1). Others indicate that, within limits, tissue receptivity is relatively independent of molecular structure and depends chiefly upon the physical status of the polymer (2). Carcinogenic reactions are elicited when the polymer is implanted as a film, and such reactions are minimal or obviated when the polymer is implanted in powder form.

Because of the uncertainty which exists in the interpretation of the results, it is important in the development of materials for internal body use that cognizance be taken of both views.

MATERIALS DESIGN CRITERIA

For a polymer to be a candidate for biomechanical application, it should be biologically receptive. This implies stability in the environment in which it is to be used as well as a physical form which does not inhibit normal cell metabolism.

Stability is a function of polymer structure, and the following attributes contribute to polymer stability.

1) The polymeric material should be chemically saturated.

2) It should not have labile groups on the side chains or chain backbone.

3) It should be vulcanizable and thermosetting, or the structure should be regularly oriented permitting close packing of the chains and crystallization.

To make certain that the polymer does not cast a "metabolic shadow," the polymer should be used in textile or porous form having interconnecting pores. To develop porous structures with adequate mechanical properties, composites consisting of fabric-reinforced porous polymers could be utilized.

In addition to biological receptivity, the polymer should have the required mechanical properties, be easily fabricated into desired shapes, and be easily sterilized.

The ranges of mechanical properties of interest are those represented by the soft elastomeric materials to flexible, but more rigid materials. For example, for vascular prostheses and patch grafts, soft elastomeric properties are desired; for femoral heads and artificial tracheae, more rigid materials appear to be required.

Elastomers and resins based on acrylic and methacrylic acid esters are chemically saturated and amenable to non-sulfur vulcanization techniques, and rigid resins synthesized from these materials have shown salutary tissue response in specific areas of the body, such as in repair of skull defects and in the mouth. It was, therefore, decided to prepare an elastomeric material based on heteropolymers of these esters. After considerable experimentation, a terpolymer elastomer was evolved which consisted of butyl acrylate, methyl methacrylate and methacrylamide. To develop adequate mechanical properties, this base elastomer was reinforced with a vinylic type filler based on the resinous polyethyl methacrylate and, to insure dimensional stability and enhanced resistance to the biological environment in which it was to be used, the elastomer was crosslinked with formaldehyde. Techniques were also devised to stiffen the base elastomer so that a range of materials of varying mechanical properties could be obtained.

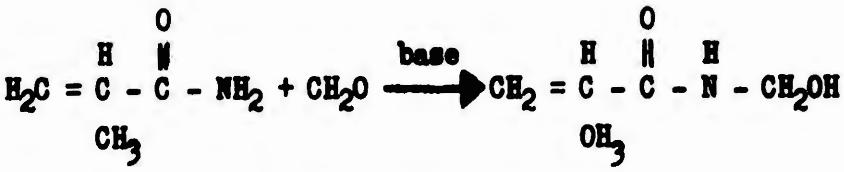
Polymer Synthesis

The terpolymer elastomer butyl acrylate (90), methyl methacrylate (7.5) and methacrylamide (2.5) was prepared and compounded with polyethyl methacrylate (37 pts/100 pts of elastomer) and formalin as described in Figure 1.

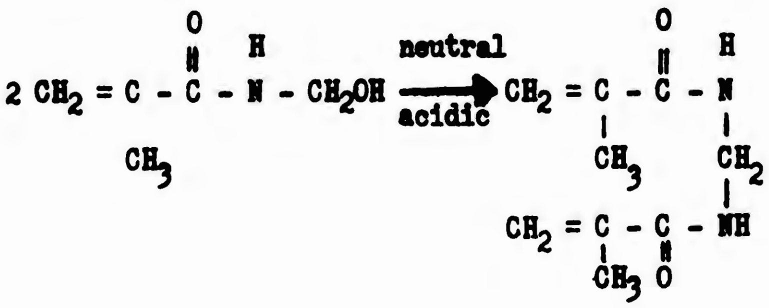
Vulcanization Reaction

The terpolymer of butyl acrylate, methyl methacrylate and methacrylamide contains amide groups which may react with formaldehyde to effect crosslinking of polymer chains. The mechanism for the reaction of methacrylamide with formaldehyde under alkaline and acid conditions may be represented as (3,4,5):

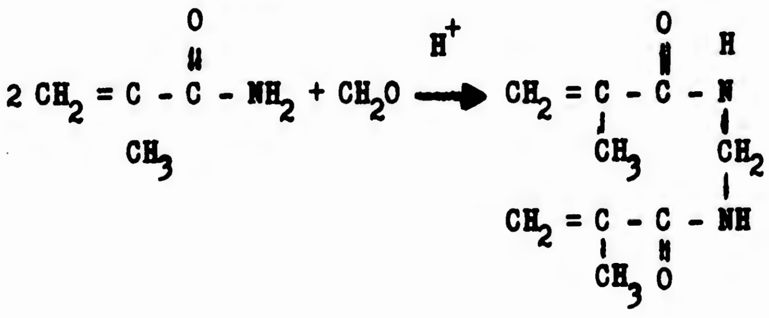
a) Alkaline



Upon conversion to neutral or acidic conditions and application of heat, water is removed and methylene diamide crosslinks are formed (3,4,5).



b) Acid



Under alkaline conditions, therefore, 1 mole of formaldehyde reacts with 1 mole of methacrylamide to form a methylol derivative. Under acidic conditions, 1 mole of formaldehyde reacts with 2 moles of methacrylamide to form methylene bis methacrylamide. Analyses of acid and alkaline latices for formaldehyde consumption are in agreement with the proposed mechanisms.

In further confirmation of the proposed crosslinking mechanism, it had been previously shown that films cast from acidic latices became more highly crosslinked as a function of latex aging at room temperatures, whereas alkaline latices did not. (6)

Typical stress-strain properties in tension of the crosslinked elastomer films as a function of formaldehyde concentration are summarized in Figure 2. Curves A, B and C were obtained at a strain rate of 0.2 in/min on films vulcanized with 0, 1 and 2 moles of formaldehyde per basal mole of methacrylamide, respectively. It is observed that as the concentration of formaldehyde is increased, the curves are shifted toward the stress axis. In the D curves, strain rate of 20 in/min, for films vulcanized with 1/2, 0, 1 and 2 moles of formaldehyde per basal mole of methacrylamide, the curves are coincident up to an elongation of 500%, beyond which they separate. However, there is no relation between formaldehyde concentration and apparent film stiffness, probably because the rate of strain is too rapid to permit the polymer chains to relax to equilibrium stress-strain values. It is of interest to note that the films appear considerably stiffer when measured in tension at a strain rate of 20 in/min.

In some internal body applications, such as femoral heads, arteries, etc., it was of interest to maintain the flexibility of the bulk polymer, but to enhance the resistance of the polymer surface to attack by body fluids. In effect, it was desired to achieve a film whose surface was highly crosslinked and hardened while maintaining the bulk of the film lightly crosslinked, soft and rubbery. To accomplish this, the films which gave stress-strain properties represented by the D curves were post dipped in a solution of formaldehyde in methanol for 15 minutes. The films were air dried for 15 minutes, oven dried at 50°C for 30 minutes, and again heated at 100°C for 1 hour.

The effects on stress-strain curves of samples so treated are illustrated in curves E through H. In all instances the curves are shifted towards the stress axis, indicating an overall stiffening of the elastomer. Those samples initially more lightly crosslinked (containing lower formaldehyde concentrations) were more greatly affected by the post treatment.

Since in all internal applications the material would require steam autoclaving for sterilizing purposes, it was of interest to determine the stress-strain properties of the elastomer using a steam cure. The results indicate that the steam cure more effectively cured the elastomer, resulting in a stiffer material. These results may be due to the hydrophilic nature of the film, which swells in steam, thereby making more reaction sites available for the crosslinking reaction with formaldehyde.

Further investigation revealed that a wide range of film-forming compositions, varying from soft elastomers to relatively stiff materials, may be prepared by varying the concentrations of terpolymer ingredients and/or by varying the concentration of the vinylic filler as well as by varying the concentration of the crosslinking agent.

Typical results are summarized in the following tables:

(For Tables I and II see next page.)

It is observed in Table I that as the concentration of butyl acrylate in the terpolymer is increased from 62.5 pts to 90 pts, the tensile stress at 300% elongation (tensile modulus) decreased, the ultimate elongation increased, and the ultimate tensile strength and hardness decreased. The flexural modulus at varying temperatures also decreased with increasing butyl acrylate concentration.

The data in Table II indicate the increase in rigidity obtainable by increasing the concentration of vinylic fillers.

In addition to the variation in mechanical properties obtainable by these means, the latex-dispersed terpolymers may be fabricated into foamed structures and a new range of properties is obtainable. A typical foam (Fig. 1, Engineering column) was prepared by adding to the stirred latex contained in a beaker the citric acid, ammonium hydroxide and sodium polyacrylate. This mixture was then whipped with a mechanical beater, first at high speed to cause the foam to rise threefold, and then at a slower speed (about half of former) to consolidate air cells into a uniform distribution. The sodium fluosilicate was then added and whipping of the foam continued at moderate speeds for 30 seconds. Whipping was stopped and the foam so prepared could be poured into flat molds for the preparation of foamed sheets or could be used to coat tubes as desired. Using carefully controlled conditions, pore sizes of 0.153 ± 0.049 mm were obtained.

TABLE I

Init. Monomer Concentration wt/% BA/MMA/MAD	Parts by wt PEMA	Tensile Strength p.s.i.	Elongation %	300% Tensile Stress p.s.i.	Tear Resistance p.i.	Shore Durometer "A"	Modulus in Flexure *					
							-40°C	-30°C	-20°C	-10°C	0°C	+10°C
62.5/35/2.5	37	1870	310	793	175	90	115000	110000	85000	55000	22000	7500
80/17.5/2.5	37	1378	448	338	90	60	50000	26000	4300	1600	850	600
85/12.5/2.5	37	1295	645	160	88	55	50000	9500	2000	1500	1000	650
90/7.5/2.5	37	1130	830	140	80	55	16000	3200	1850	900	700	575

* A.S.T.M. D 747-50

TABLE II

BA/MMA/MAD	Parts PEMA	Tensile Strength	Per Cent Elongation	100% Tensile Stress
90/7.5/2.5	20	1130	830	140
"	30	1235	755	265
"	40	1430	705	340
"	50	1580	590	505

Abbreviations key: BA--Butyl Acrylate; MAD--Methacrylamide; MMA--Methyl Methacrylate; PEMA--Poly Ethyl Methacrylate

Applications

Many surgical repair applications for these materials having a wide range of mechanical properties were suggested. These include vascular prostheses such as artificial arteries and patch grafts, caps for replacement of diseased cartilage on femoral heads, tracheae, heart valves, bile ducts, and non-adherent surgical bandages for wounds and burns. Of these, the vascular prostheses (Fig. 1) which have been evaluated most extensively by surgeons, and the cartilaginous replacements shown by the femoral caps in Figure 3 are examples of composite structures utilizing soft, rubbery compositions and more rigid compositions, respectively. The preparation of these two prostheses is described to illustrate the fabrication procedures developed.

Artificial Artery. A typical artificial artery is a composite structure consisting of a crimped dacron tube which is coated with the latex foam. The coating apparatus is shown in Figure 4. The apparatus consists of a polypropylene container whose bottom has been replaced with a piece of 5 mil polyethylene film through which holes have been made to accommodate the guide element of the applicator. The polyethylene frame (c) is tubular in shape with large apertures to permit easy invasion by the foam. Into the ends of this tubular frame is inserted an element consisting of a guide (d) which stabilizes the assembly against lateral movement, a guard (e) which prevents the applicator from slipping back through the holes of the thin polyethylene film, and a compressor element (f) which is a polyethylene ring whose purpose is to bring the foam into better contact with the fabric by lightly ironing the latter. The applicator element is concentric with the circumference of the mandrel, which provides an annular space through which the foam has to pass. This element regulates the thickness of the foam deposit.

The fabric utilized has been knitted dacron Fluflon ranging from 140 denier to 40 denier and in various thread counts from 77 counts per inch to 25 counts per inch. The type of fabric for the arterial graft was selected for use by its anatomical location. The heavier fabrics were used where the blood pressures were highest. The optimum thread count still remains to be determined.

In a typical coating experiment, the crimped dacron tubes were first soaked in a dilute (6% solids) terpolymer latex for 1 hour. They were removed and permitted to oven dry at 50°C for a few minutes, and then stored in a humidity cabinet at 100% relative humidity.

The impregnated tubes were placed over glass mandrels, stretched 80% to 90% over their original length, and held in this position by tape. The polypropylene container was filled with the compounded foam and the impregnated tubes were drawn through the foam in the coating apparatus to their full length, slowly but continuously, taking about 20 seconds for the whole operation. The tubes were then dried for 2 hours at 60°C in a circulating-air oven.

After drying the tubes, they were placed in a cylinder containing cold water and the cylinder was immersed in a hot water bath at 100°C for 30 minutes. The tubes were removed and allowed to cool to room temperature. After removing the tape from one end of the coated dacron tube, the crimp was re-established by carefully pushing the loose end toward the taped end. Great care was taken during this step to prevent undue compression. The tape was re-applied to prevent the tube from slipping and the assembly was again placed in hot water (90°-100°C) for 30 minutes to heat-set the foam to the newly crimped shape. The tape was removed and the coated dacron tube pushed off the glass mandrel into a water bath. Water-soluble compounding ingredients were eluted by placing the tube in a circulating water bath for 24 hours. Further cure was accomplished by heating the tube in dilute (5%) aqueous formalin for 1 hour followed by steam vulcanization at 250°F for 1 hour.

Cartilaginous Replacement for Femoral Head. Starting with a dog's original femoral head to obtain the general shape, a one-piece negative mold was formed from silicone rubber. A positive dipping form was subsequently made using epoxy resin.

The following dipping procedure was used:

Dacron fabric was shaped over the epoxy model of the femoral head and secured with string at the femoral neck. Fabric wrinkles were removed wherever they appeared on the surface in order to provide the smoothest surface possible.

The epoxy mold, with fabric attached, was dipped into latex to saturate the cloth. The assembly was then immersed in concentrated coagulant (30% calcium nitrate in denatured alcohol), followed by dipping in latex for 3 minutes. The coagulated film was leached in warm water for 1 hour and then dried for several hours in an oven at 60°C until clear film resulted. The composite was post treated for 1 hour in dilute boiling formalin solution, then autoclaved at approximately 258°F for 30 minutes and finally dried in an oven at 60°C.

The composite was allowed to cool to room temperature and then stripped from the form. Approximate thickness of the composite was 1 mm.

The latex system was composed of butyl acrylate, methyl methacrylate and methacrylamide (80/17.5/2.5), compounded with 37 parts of polyethyl methacrylate and 2 moles of formaldehyde per basal mole of methacrylamide.

SUMMARY

Biologically receptive terpolymers based upon acrylic and methacrylic acid esters and methacrylamide were prepared and their properties described. By making variations in terpolymer compositions, formaldehyde concentration and vinylic filler loading, it was possible to obtain qualitatively similar materials with a wide range of mechanical properties. A method for preparing the materials in sponge form was developed.

The application of the materials to the preparation of vascular prostheses and to a femoral head cap is described.

LITERATURE CITED

(1) Hueper, W.C. A.M.A. Archives of Pathology, June 1959, Vol. 67, pp. 589-617.

(2) Oppenheimer, et al. Cancer Res. 15: 333-340, 1955.

(3) Brit. Patent 467,492 (June 11, 1937).

(4) U.S. Patent 2,173,005 (Sept. 12, 1939).

(5) Brit. Patent 482,897 (April 7, 1938).

(6) Leonard, F., Nelson, J., and Brandes, G. Ind. Eng. Chem. 50, 1053-8 (1958).

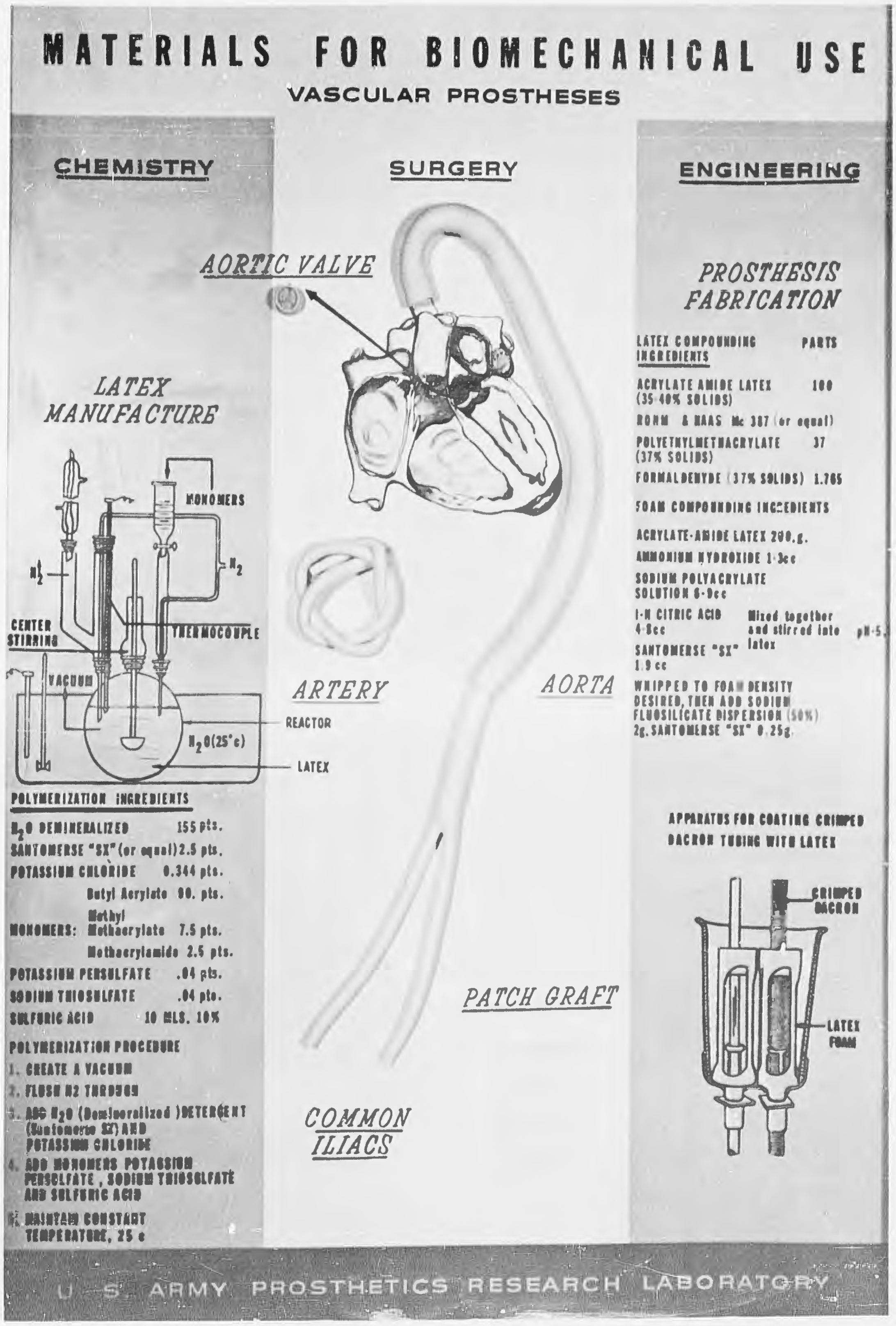

MATERIALS FOR BIOMECHANICAL USE
VASCULAR PROSTHESES
CHEMISTRY
SURGERY
ENGINEERING
AORTIC VALVE
LATEX MANUFACTURE
MONOMERS
N_2
N_2
CENTER STIRRING
THERMOCOUPLE
VACUUM
H_2O(25°c)
REACTOR
LATEX
POLYMERIZATION INGREDIENTS
H_2O DEMINERALIZED 155 pts.
SANTOMERSE "SX" (or equal) 2.5 pts.
POTASSIUM CHLORIDE 0.344 pts.
MONOMERS: Butyl Acrylate 90. pts.
Methyl Methacrylate 7.5 pts.
Methacrylamide 2.5 pts.
POTASSIUM PERSULFATE .04 pts.
SODIUM THIOSULFATE .04 pts.
SULFURIC ACID 10 MLS. 10%
POLYMERIZATION PROCEDURE
1. CREATE A VACUUM
2. FLUSH N2 THROUGH
3. ADD H_2O (Demineralized) DETERGENT (Santomerse SX) AND POTASSIUM CHLORIDE
4. ADD MONOMERS POTASSIUM PERSULFATE, SODIUM THIOSULFATE AND SULFURIC ACID
5. MAINTAIN CONSTANT TEMPERATURE, 25 c
ARTERY
AORTA
PATCH GRAFT
COMMON ILIACS
PROSTHESIS FABRICATION
LATEX COMPOUNDING INGREDIENTS PARTS
ACRYLATE AMIDE LATEX (35-40% SOLIDS) 100
ROHM & HAAS Mc 387 (or equal)
POLYETHYLMETHACRYLATE (37% SOLIDS) 37
FORMALDEHYDE (37% SOLIDS) 1.765
FOAM COMPOUNDING INGREDIENTS
ACRYLATE-AMIDE LATEX 200.g.
AMMONIUM HYDROXIDE 1-3cc
SODIUM POLYACRYLATE SOLUTION 6-9cc
1-N CITRIC ACID 4-8cc
SANTOMERSE "SX" 1.9 cc
Mixed together and stirred into latex
pH-5.
WHIPPED TO FOAM DENSITY DESIRED, THEN ADD SODIUM FLUOSILICATE DISPERSION (50%) 2g. SANTOMERSE "SX" 0.25g.
APPARATUS FOR COATING CRIMPED DACRON TUBING WITH LATEX
CRIMPED DACRON
LATEX FOAM
U S ARMY PROSTHETICS RESEARCH LABORATORY

Fig. 1

STRESS STRAIN CURVES
FOR CROSS LINKED TERPOLYMER ELASTOMER

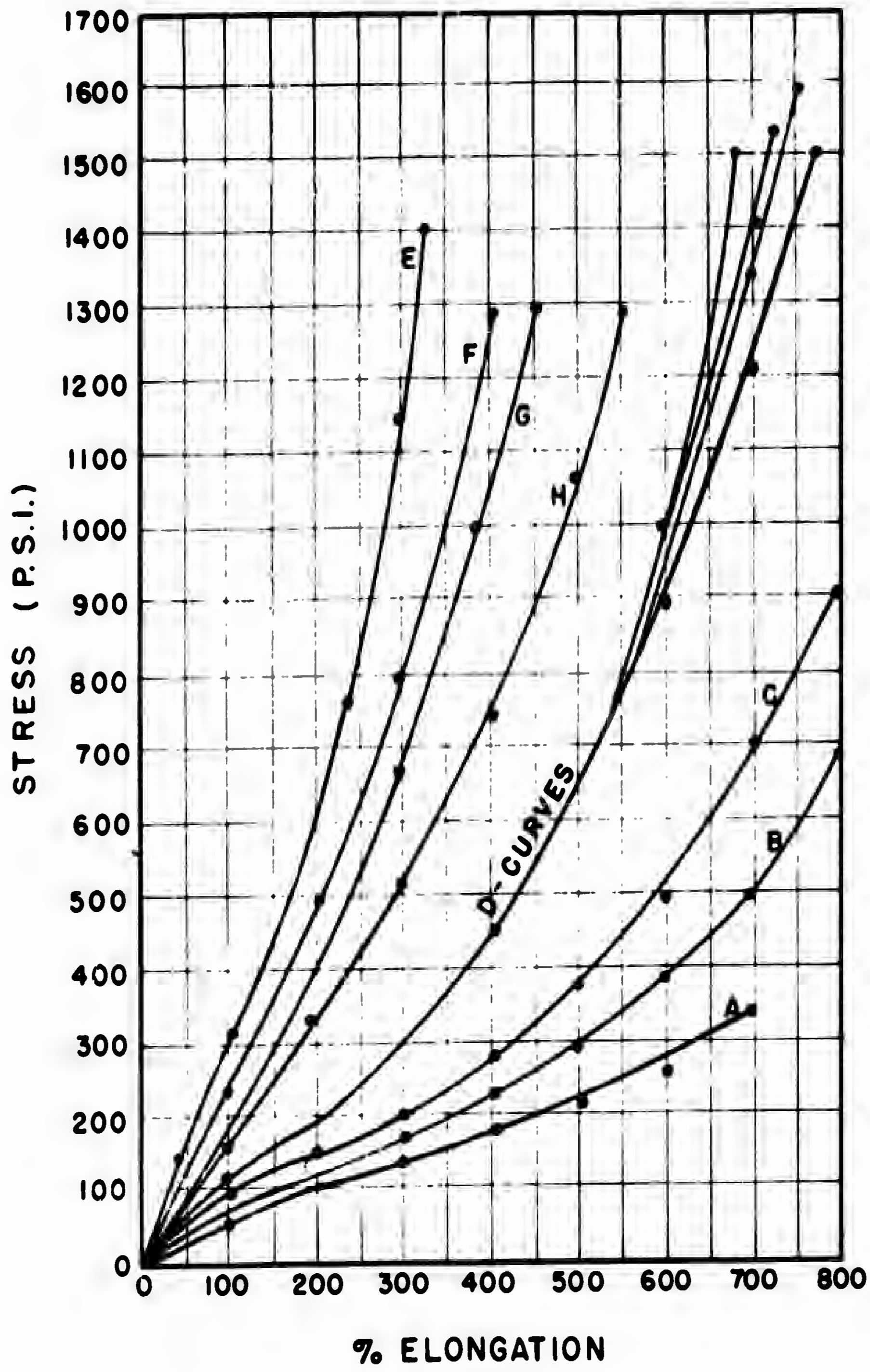
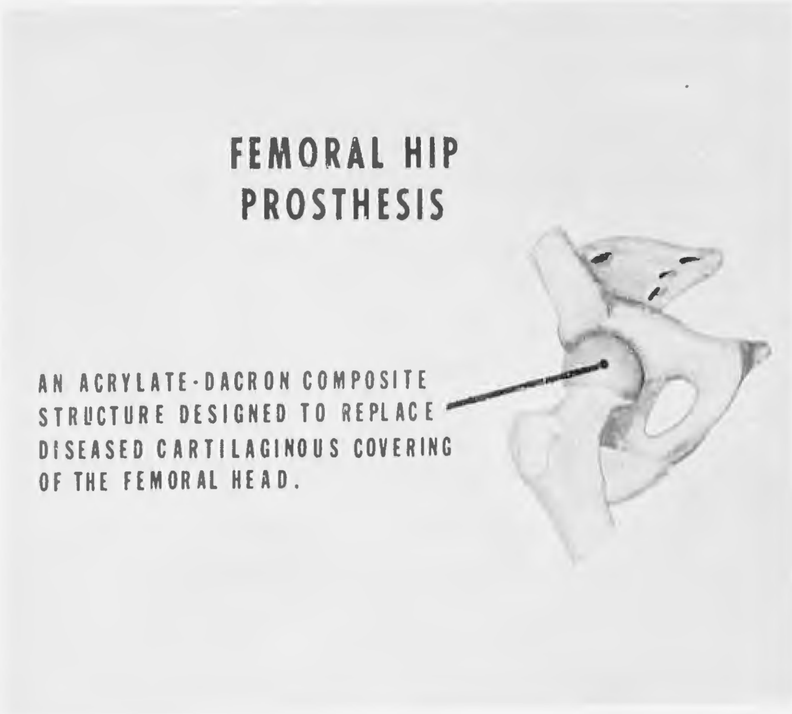


Fig.2


FEMORAL HIP
PROSTHESIS
AN ACRYLATE-DACRON COMPOSITE
STRUCTURE DESIGNED TO REPLACE
DISEASED CARTILAGINOUS COVERING
OF THE FEMORAL HEAD.

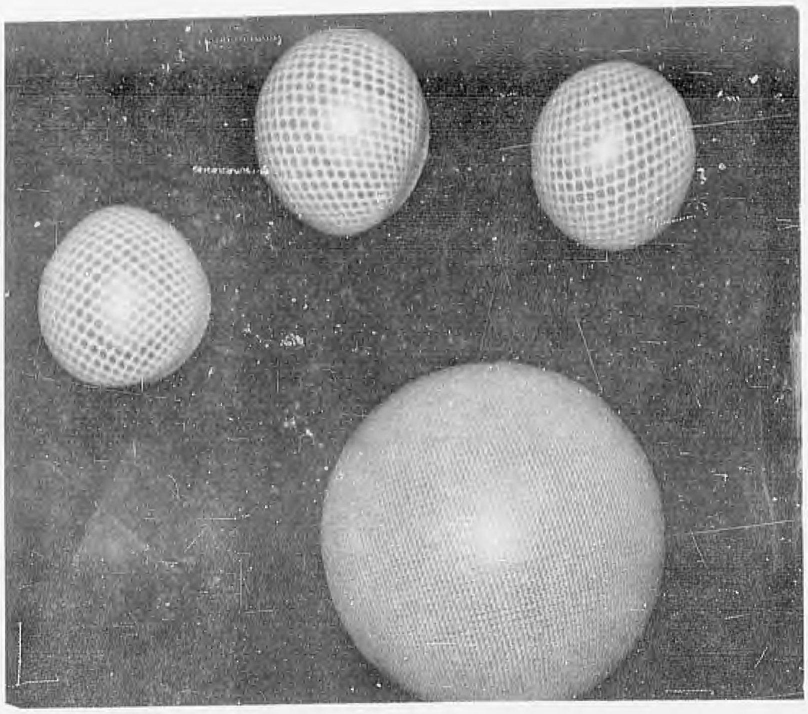


Fig. 3

APPARATUS FOR COATING CRIMPED DACRON TUBING WITH LATEX FOAM

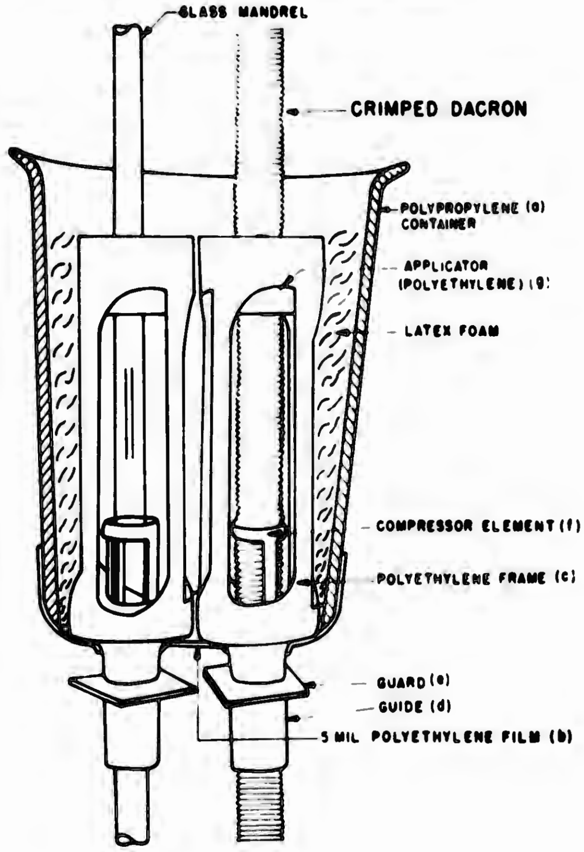


Fig. 4

A NOVEL CARBOHYDRATE REACTION

LOUIS LONG, Jr., DEREK H. BALL, and A. K. MITRA
QUARTERMASTER RESEARCH & ENGINEERING COMMAND
NATICK, MASSACHUSETTS

A study of the possibility of making polymers from sugar, the everyday table variety, was begun in 1955. There were several attractive reasons for doing so. Sugar, chemically known as sucrose, has the largest production of any pure organic chemical. It is obtained by refining cane sugar to a purity of 99.96 per cent. It is also one of the cheapest materials we have, selling at a few cents per pound in the wholesale market. Not only is there an annual overproduction of sugar, but there exists a potential overproduction of at least twice the present annual tonnage.

Economically, therefore, sugar is an attractive raw material, because of its almost limitless availability and its stable, low price. The supplies of coal and petroleum, which now form the basis of most synthetic organic chemicals, will some day be depleted, and we shall be forced to depend upon atomic energy and the renewable sources of organic compounds produced directly by vegetation.

Because of its favorable position economically, it seemed to offer excellent possibilities for use as a raw material for polymers. These might be used as plastics in the form of sheets, rods and other shapes, or as adhesives to bind together wood, metals or, particularly, glass. It was suggested that a plastic wrapper, or foam material, made of a sucrose polymer might be edible, thereby diminishing the weight of materiel that a soldier would have to carry in the field.

As a result of the work described subsequently and related experiments, two patents have been applied for on behalf of the United States Government-- one for polymeric materials made from sugar and the other for adhesives similarly prepared. Thus, in part, our original purpose has been achieved.

To form a polymer from sucrose, it was planned to make an elongated molecule with two reactive ends from which polymeric materials could be made by condensation with other bifunctional molecules, similar to the compounds used to form nylon.

The structure of the sucrose molecule has been demonstrated unequivocally by X-ray diffraction (1). Its synthesis was accomplished by R. U. Lemieux in 1953 (2). As shown in the first slide, it has three primary hydroxyl groups and five secondary ones. From the work of previous investigators (3), it has been known for some years that primary hydroxyls are more readily tosylated and mesylated than secondary ones (4). It was expected, therefore, that a tri-tosyl or tri-mesyl monomer would be readily obtained. (Slide Two) Since it was also known that the one prime position of sucrose was less reactive than the 6 and 6' positions, it was clear that such a molecule would resemble the desired model of an elongated compound with two reactive ends (5, 6).

Based on these premises, experimental work was begun, with the result that amorphous materials were obtained which gave empirical analyses corresponding to tetramesyl and tetratosyl sucrose derivatives. This research was supported by the Sugar Research Foundation of New York, initially, and, later, by the National Science Foundation and the National Academy of Sciences. (Slide Three) The fourth mesyl or tosyl group was considered to be on the two-carbon atom by analogy to a considerable volume of supporting evidence in the literature (3). To prove the structure of these tetra-substituted sucrose derivatives, the following reactions were suggested as offering a feasible method. (Slide Four) After methylation and de-mesylation or de-tosylation, it was planned to hydrolyze the tetramethyl compound so obtained and to isolate and identify 3,4-di-O-methyl-D-glucose and 3,4-di-O-methyl-D-fructose. The proof of the presence of these compounds in the hydrolyzate in good yield would afford unambiguous evidence for the structure of the tetra-substituted sucrose derivatives.

Since 3,4-di-O-methyl-glucose would be needed for the proof of the positions of substitution of the mesyl groups in tetramesyl-sucrose, its synthesis was attempted. As the previously known synthesis of it by Bell and Greville (7) required nine steps with a very low over-all yield, other methods of preparation were investigated. In one of these (Slide Five) the mesylation of methyl α-D-glucopyranoside was examined. Helferich and Gnüchtel (8) and, more recently, Cramer et al. (9), treated methyl α-D-glucopyranoside with one molar equivalent of mesyl chloride at -20°. No crystalline material was isolated from the reaction product but, after acetylation, crystalline methyl 2,3,4-tri-O-acetyl-6-O-mesyl-α-D-glucoside was obtained in good yield (Cramer et al. (9) reported 67%). The reaction conditions described by these authors were followed and the mesylation product was examined by paper chromatography. The chromatograms indicated that, in addition to a mono-esterified derivative, considerable

amounts of methyl α-D-glucopyranoside and of di- and tri-esterified derivatives were present. The relative amounts of these constituents were not appreciably affected by lowering the reaction temperature to -40°. The molar ratio of mesyl chloride to glucoside was increased and when 2.2 or more, no starting material or mono-substituted products were observed. The product appeared to be mainly a di-O-mesyl glucoside together with smaller amounts of higher substituted glucosides.

From the reaction of methyl α-D-glucopyranoside with 2.2 moles of mesyl chloride, a crystalline methyl di-O-mesyl-α-D-glucoside was isolated in 51% yield. At this point, Dr. Mitra returned to India, and the work was continued by Dr. D. H. Ball, supported by a grant from the Office of the Surgeon General. A new crystalline compound had been synthesized, but its structure had not been proved. By analogy with previous work (10, 11), the mesyl groups were thought to be at C_2 and C_6 of the glucopyranoside and this was confirmed by methylation followed by alkaline saponification and acid hydrolysis to 3, 4-di-O-methyl-D-glucose. Methylation of the methyl di-O-mesyl-α-D-glucoside by the procedure of Kuhn (12) afforded crystalline methyl 2, 6-di-O-mesyl-3, 4-di-O-methyl-α-D-glucoside in 91% yield. Although many attempts to remove the mesyl groups by the use of sodium amalgam reduction according to the method of Freudenberg were made, this reaction could not be accomplished. However, treatment of it with boiling aqueous sodium hydroxide removed the mesyl groups and gave, after distillation, methyl 3, 4-di-O-methyl-α-D-glucoside, a new compound, in 64% yield. Acid hydrolysis of the latter yielded 3, 4-di-O-methyl-D-glucose which was further characterized as the crystalline anilide. The overall yield of 3, 4-di-O-methyl-α-D-glucose from methyl α-D-glucopyranoside was approximately 25%; a considerable improvement on previous syntheses (7, 13).

Acetylation and benzoylation of methyl 2, 6-di-O-mesyl-α-D-glucoside gave products which failed to crystallize, but, with fuming nitric acid in acetic anhydride, the crystalline di-nitrate was formed.

Although the synthesis of 3, 4-di-O-methyl-α-D-glucose had been accomplished most successfully, we were not satisfied to discontinue these experiments. We had observed certain anomalies in the course of the saponification of the mesyl groups and wished to find out what was happening. As a result of this scientific curiosity, a novel carbohydrate reaction was discovered which we hope to investigate further in the future. (Slide Six)

When aqueous methanol was used as the solvent in the saponification of methyl 2, 6-di-O-mesyl-3, 4-di-O-methyl-α-D-glucoside, the sulfur-free product contained two components. After acid hydrolysis, two spots were obtained on paper chromatograms,

the slower moving corresponding to 3, 4-di-O-methyl-D-glucose. The faster moving component had the mobility expected for a tri-O-methyl hexose. When methyl 2, 6-di-O-mesyl-3, 4-di-O-methyl-α-D-glucoside was treated with sodium methoxide in anhydrous methanol, methyl 3, 4-di-O-methyl-α-D-glucoside was not formed and the sulfur-free product after distillation analyzed for a methyl tri-O-methyl hexoside. This compound crystallized when cooled to 0^{o} and acid hydrolysis afforded 3, 4, 6-tri-O-methyl-D-glucose, proving that the methyl tri-O-methyl hexoside is methyl 3, 4, 6-tri-O-methyl-α-D-glucoside. Treatment of this glucoside with mesyl chloride in pyridine gave methyl 2-O-mesyl-3, 4, 6-tri-O-methyl-α-D-glucopyranoside.

The direct nucleophilic displacement of a primary mesyloxy group by methoxide ion does not appear to have been observed previously in the carbohydrate field, although displacements have been effected by a wide variety of anions (14). Taylor and Kent (15) found that when 1, 2:3, 4-di-O-isopropylidene-6-O-mesyl-α-D-galactose was treated with potassium fluoride in methanol at 150^{o}, the products contained methoxyl groups as well as the anticipated fluorine and when 1, 6-anhydro-3, 4-O-isopropylidene-2-O-mesyl-β-D-galactose was treated with these reagents, the axial C_2-mesyloxy group was displaced, with inversion, by methoxide. The product did not contain fluorine and was characterized as 1, 6-anhydro-3, 4-O-isopropylidene-2-O-methyl-β-D-talose (16).

The comparatively mild conditions under which we observed displacement at C_6 may indicate that the sulfonyl group at C_2 exerts an effect and it is hoped that future work will clarify this point.

In preliminary tests against the Dunning leukemia tumor, carried out under the auspices of the National Institutes of Health, methyl 2, 6-di-O-mesyl-α-D-glucoside showed activity, and additional quantities are currently being examined. This compound is also a key substance in the synthesis of a potential radiation protective material.

REFERENCES

(1) C. A. Beevers and W. Cochran, Proc. Roy. Soc. (London), A190, 257 (1947).

(2) R. U. Lemieux and G. Huber, J. Am. Chem. Soc., 75, 4118 (1953).

(3) J. M. Sugihara, Advances in Carbohydrate Chem., 8, 1 (1953)

(4) "Tosyl" = p-Tolylsulfonyl; "Mesyl" = Methylsulfonyl

(5) R. C. Hockett and Morris Zief, J. Am. Chem. Soc., 72, 1839 (1950).

(6) A. L. Raymond and E. F. Schroeder, U. S. Patent, 2, 365, 776 (1944).

(7) D. J. Bell and G. D. Greville, J. Chem. Soc., 1902 (1950).

(8) B. Helferich and A. Gnüchtel, Ber., 71, 712 (1938).

(9) F. Cramer, H. Otterbach, and H. Springmann, Chem. Ber., 92, 384 (1959).

(10) T. Lieser and R. Schweizer, Ann., 519, 271 (1935); Naturwissenschaften, 23, 131 (1935).

(11) J. Asselineau, Bull. soc. chim. France, 937 (1955).

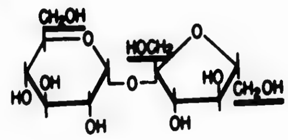
(12) R. Kuhn, H. Trischmann, and I. Löw, Angew. Chem., 67, 32 (1955).

(13) J. Dewar and G. Fort, J. Chem. Soc., 496 (1944).

(14) R. S. Tipson, Advances in Carbohydrate Chem., 8, 107 (1953).

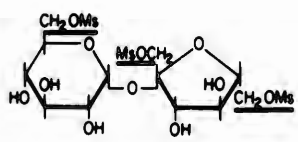
(15) N. F. Taylor and P. W. Kent, J. Chem. Soc., 872 (1958).

(16) P. W. Kent, D. W. A. Farmer, and N. F. Taylor, Proc. Chem. Soc., 187 (1959).



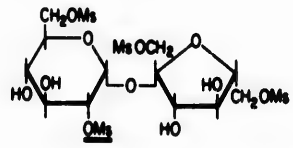
SUCROSE

SLIDE 1



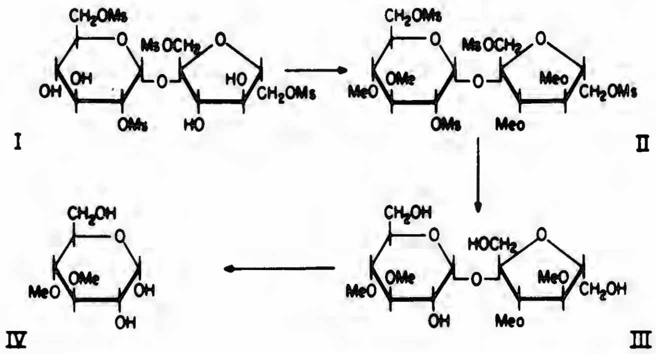
TRI-MESYL SUCROSE

SLIDE 2



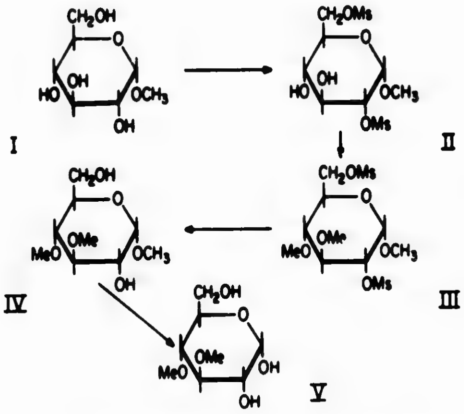
TETRA–MESYL SUCROSE

SLIDE 3



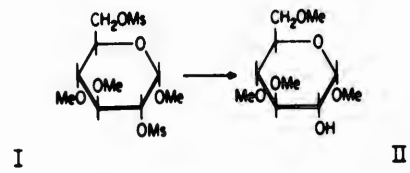
TETRA–MESYL SUCROSE PROOF OF STRUCTURE

SLIDE 4



SYNTHESIS OF 3,4-DI-O-METHYL-GLUCOSE

SLIDE 5



A NOVEL CARBOHYDRATE REACTION

SLIDE 6

WINDS AND ATMOSPHERIC TURBULENCE AT VERY HIGH ALTITUDES

MARVIN J. LOWENTHAL
U. S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY
FORT MONMOUTH, NEW JERSEY

Prior to 1955, knowledge of the wind field was limited to the lowest 20 km of the atmosphere. Information was obtained from sounding balloons whose bursting altitudes were generally below 25 km. Occasional ascents to 30 km were achieved, mainly in daylight hours, and under light wind conditions. The IGY period (1957-58) provided the necessary impetus for improvement. With new balloons, developed by the U. S. Army Signal Corps, routine flights to 30 km became possible, so that daily 10 mb maps (approx 31 km) are now constructed on a hemispheric basis.

Expanded requirements necessitated by the ever increasing use of missiles and rockets, clearly showed that balloon-borne sounding equipment must be replaced by rocket-sondes. Once more the U. S. Army Signal Corps led in the establishment of a meteorological rocket network to measure winds up to 60-70 km. Starting in the fall of 1959 with three stations, the network has gradually expanded to its current eight stations. Winds are determined from radar tracking of chaff, parachutes, or plastic balloons, ejected from small meteorological rockets.

At first glance it would appear that this method could be used for all altitudes, but this is not the case. Above 70 km, the parachute, plastic balloon, and even the lightest chaff fall with tremendous speeds. Their sensitivity, therefore, is so low that no accurate wind measurements can be made. Thus, 70-75 km is almost the limit of their usefulness. Even with this limitation, use of rocket measured winds has enabled us to depict the mean wind flow at 1 mb or 0.5 mb, (50-60 km) more than doubling the former altitudes reached by balloon flights.

For wind measurements above 70 km, several methods have been tried with varying degrees of success. Among these are radio echoes from meteor trails, (80-110 km), Rocket Grenade Experiments

(30-90 km), ejection of alkali vapor at twilight (80-200 km) and on-board instrumentation on large rockets. Many of these techniques are complex and require extensive instrumentation. A much simpler method without an all weather capability, is the measurement of winds using Rocket Exhaust Trails.

The condensation trail of high altitude aircraft is a familiar sight and the deformations of the trail due to the wind is equally as well known. If photographs of the trail are taken at short intervals, the magnitudes of the winds and wind shears can easily be computed. Similarly, if pictures of the exhaust trail of a rocket can be taken, a knowledge of the winds and their behaviors at altitudes of 100 or even 200 km can be gained, the altitudes being dependent on the height of the rocket at burnout. The information contained here was obtained from stereoscopic photographs of three exhaust trails of the second-stage rocket of Project Shotput, a NASA project at Wallops Island, Virginia. The rockets were fired as tests for the ejection of the 100 foot plastic balloon, that ultimately became the Echo Satellite. A two stage rocket was used for the tests, the second stage ignition occuring about 60 km, with burnout between 100 and 120 km. The tests were carried out at twilight after sunset had occurred at the surface, but the upper levels of the atmosphere were still illuminated. This period of the day permits observers to see the ejection of the balloon and its inflation without interference from the normal daylight glare. This is also the best time to photograph the exhaust trail, since the maximum contrast between trail and background is obtained. In the absence of clouds or visibility restrictions the trail can be observed several hundred kilometers up and down the eastern seaboard.

In order to determine both horizontal components of the wind (N-S & E-W), stereoscopic photographs were made. One camera was located at the USASRDL, Belmar, N. J., the other at Brookhaven Laboratories, Long Island. Pictures were made every 15 seconds at Brookhaven and every 30 seconds at Belmar. Evaluation was performed by the U. S. Army, Engineering R&D Laboratory, Ft. Belvoir, Virginia.

Three different trails were analyzed (1) 27 February 1960 (2) 1 April 1960, and (3) 31 May 1960. There will be discussed in detail.

1. Analysis of exhaust trail on 27 February:

At 1825 EST on 27 February 1960 the first of the photographs was made. Figure 1 shows the exhaust trail a few seconds after its formation. Already slight deformation of the original straight line path has taken place in the lower portion, and a discontinuity appears near the top. The density of the trail above the discontinuity is much less dense, as if the rocket engine had been cut off, and the last drops of fuel were being expended. It is by

no means sure that this is the correct explanation, since subsequent events indicate a discontinuity in wind structure at this level. It may well be that the changed appearance is due to the environment.

Figure 2, taken two minutes later, shows the tremendous wind shears acting on the trail. The camera faced almost due south hence as displacement to the right of the original position was caused by an east wind (blowing from East to West); to the left, by a west wind. The quantitative evaluation of the winds at the various levels are listed in Table I below:

TABLE I
Wallop's Island, Va. Winds on 27 Feb 1960 1825 EST

Point	Altitude(km)	Wind Speed mps	Wind Direction(degrees)
A	64	76	300
B	71	69	280
C	74.5	93	255
D	76	13	155
E	79	29	225
F	83	36	210
G	96	198	245
H	101	136	245
I	104	21	360
J	105	39	205
K	109	106	140
L	116	Too Faint for Evaluation	

The wind profile obtained from Table I shows a region of strong winds at 96 km from the southwest (245^{0}) of nearly 200 meters per second. Only 13 kms higher, there exists a region of strong easterly winds with speeds greater than 100 meters per second. The region between 96 & 109 km is characterized by large wind shears. In particular, at 104 km, the shear was greater than 70 meters/sec/km. This means that a vector wind change of 70 meters/sec occurred with a change of 1 km in altitude.

Another singularity was visible in the lower portion of the trail. A jet existed near 75 km (point C) with wind speeds of 93 mps. Shears in this region are also of the order of 70 m/sec/km over narrow zones.

2. Winds on 1 April 1960:

Figure 3 shows the photograph taken a few seconds after formation and Figure 4, 2 minutes after the formation at approx. 1858 EST. A pattern of east wind zones sandwiched between westerlies was again observed. A tabular list of values appears in Table II.

TABLE II
Wallops Island, Va. Winds on 1 April 1960

Point	Altitude(km)	Wind Speed mps	Wind Direction(degrees)
A	69.5	58	45
B	73.5	47	305
C	76	24	5
C_1	76.5	51	285
D	83	11	45
E	90	80	120
F	94	80	230
G	98	44	210
H	99	84	290
I	103	27	55
J	107	140	300
K	112	124	310
L	115	250	300

The wind profile between 69.5 km and 83 km on this day was much more complex than that of the 27th of February. Here alternate regions of easterly and westerly winds existed in contrast to the predominant westerly flow earlier. Above 90 kms the April winds are generally from the west - easterlies appearing only between 100 km and 106 km. It is noteworthy that the maximum shear occurred near 100 km, with shears greater than 85m/sec/km - almost exactly in the maximum shear region of the February shot. Another point of similarity is the sharp needle-like wind singularity at 76.5 km (C_1). This compares with point C in Figure 2.

The most important facts gathered from the two analyses is the knowledge that the strong winter west winds from 65 to 90 km have broken down. The early spring trail indicates an indifferent circulation pattern, a transitional period between the winter pattern and the summer - which will be illustrated in the next exhaust trail analysis.

3. Winds on 31 May 1960:

Poor visibility on the 31st of May precluded stereoscopic pictures. A faint trail was obtained on a ballistic camera at Belmar, but the quality of the picture was too poor for quantitative analysis. At Ft. Belvoir, Va. pictures were obtained. Unfortunately the presence of clouds made complete analysis impossible. The portion of the trail from 90 km to 104 km was evaluated, with the results shown in Table III. Because of the many difficulties, the accuracy of this analysis is much less than the others. Winds are probably accurate only to plus or minus 10 mps. Altitudes are believed accurate to 0.5 km although the entire trail may be displaced some 2-5 kms up or down.

TABLE III
Wallops Island, Va. Winds on 31 May 1961 (1952 EST)

Altitude(km)	Wind Speed mps	Wind Direction(degrees)
90	58	16
94.5	98	58
96.5	90	71
97.5	67	85
98.5	91	56
100	46	142
102.5	24	30

It should be noted that all the winds are blowing from the East - a complete reversal of direction from the winter. Again a singularity appeared near the 100 km level where the wind swings rapidly from 56° (ENE) to 142° (SSE) then back to 30° (NNE). Another similarity can be noted in all three profiles. A distinct minimum in the wind field occurs near 103 km, just above the region of maximum shear.

Are these data compatable with those of other experimenters? While few results are available, there is agreement. Manring, et alii, (1) determined winds from the ejection of alkali vapors during twilight. They obtained wind values between 77 and 110 km on 12 October 1955 and 11 April 1956. At 105 km, they obtained strong westerly winds (276°, 180 mps) with large wind shears (110 mps/km at 108 km) Ainsworth, et alii (2) obtained winds from on-board rocket instrumentation. Winds near 118 km, of 250 mps were deduced, with the maximum wind shear of 90 m/sec/km at 112.5 km. Another sodium vapor ejection made in the Sahara Desert on 10 March 1959 under the direction of Prof. J. Blamont (3) discovered shears of at least 110 m/sec/km at 110 km. It would appear, then, that a region of maximum shear exists in the vicinity of 100 km at many localities and is, apparently, a permanent part of the atmospheric circulation at high altitudes.

Additional information may be obtained from study of the trails such as vertical, motions and diffusion rates, which are beyond the scope of this paper. Turbulence, however, can be inferred from a close examination. The composition of the trail (Fig 2) is not constant throughout its length. From points F to J small whorls may be observed. It is evident that turbulent eddies exist throughout the region between 85 km and 105 km, whose diameters are roughly 1 km. The needle like spikes mentioned earlier are in reality large circular eddies, again approximately 1 km in size, evidence of extreme turbulence in that region.

These observations are qualitative only - for quantitative measurements of the variability of the winds, we must resort to other means. While no repetitive rocket trail experiments have been made, meteorological rocket firings, separated by short time intervals are

available. These are of course, at much lower altitudes, but they do represent the only data available on wind variations in the mesosphere, and will be considered briefly.

On nine different days during July and August 1960 two or more Loki II rockets were fired at Cape Canaveral at various intervals from 17 to 75 minutes. On seven of these days, two rockets were fired; one day, three rockets; and once, 4 rockets. Nylon chaff was ejected from the rocket and tracked with radar. The winds were compiled from the original radar traces. Due to the finite thickness of the trace, wind directions are valid only to the nearest 5°. The uncertainties of matching time checks to the radar track limits speed accuracies to the nearest whole meter per second. Within these limitations, wind variability may be examined.

If all firings contained data from all levels, there would be a maximum number of 12 sets of wind differences at each 5000-foot level from 80,000 feet to 210,000 feet. However, this was not the case. For example, six winds are available at 210,000 feet; and of these, only two are consecutive, giving only one difference to use for variability. This, of course, is insufficient for any analysis. To increase the reliability of the results, grouping the data is the next step. The data for three consecutive levels are lumped together, the results being plotted at the midpoint. The results, as shown in Fig 5, are meaningful. There are few wild swings to the curve, and a recognizable trend has emerged. Additional smoothing (20,000 foot layers) helps little, showing that the point of diminishing returns has been reached from the standpoint of efficiency.

From the curve, the variability is dependent on height, the mean difference in wind speed increasing from 2 mps at 90,000 feet to 8 mps at 210,000 feet. These values are somewhat higher than would be expected from studies of variability at lower levels, and the strong dependency on height has not been observed before. The phenomenon needs further investigation to discover the causes, some of which can be enunciated at this time.

It is in order to inquire whether the increasing variability with height is connected with increasing wind speeds. A plot of average wind speed against height is shown in Fig 6. It is plainly seen that the maximum wind is found at 50 km, tapering off sharply above that level. Clearly, the increased variability at the highest levels is not due to increased wind speeds, although the decreased slope of the curve with height suggests that variability is partially dependent on wind speed.

Another thought that comes to mind is the increased error in radar tracking of the chaff at the higher altitudes. This can be dispensed with in rather short order. The cloud of chaff on ejection is quite small; and the radar set, after acquisition, should be

capable of giving the position with good accuracy. As the chaff falls, it spreads into a cloud that may cover an area greater than a square mile. The radar set tracks the cloud, but does not necessarily track the same point in the cloud. Hence, some of the variability at lower levels may be attributed to tracking errors.

Tidal effects is another source of variability of winds that becomes sizable at the upper levels. It would be a factor if periods of several hours were considered, but is not sizable in the short periods being considered here.

The next question that arises is: Does the wind direction vary in the same manner? Fig. 7 shows the change in wind direction for 15,000 and 20,000 foot layers. The smoothing process is, of course, evident as the thickness of the zone is increased. The curve is somewhat different from that of wind speed. There is little difference in direction variability from 25 km to almost 50 km, with large changes thereafter. Fig. 8 shows the winds during the period studied. From 25 km to 50 km there exists a deep persistent layer of easterly winds with a slight southerly component. We may hypothesize that Cape Canaveral lies on the western edge of the general low-pressure area centered nearer the equator. With increasing height above 50 km, the center of the low moves northeastwards. A slight change in altitude changes the direction from southeast to south or even southwest. Above 60 km, the low center is definitely north of the station, with westerly or even northwesterly winds. A slight error in altitude will give rise to a great direction difference that would not obtain at lower elevations. It is believed that the large direction differences at the highest elevations are attributable, in part, to this uncertainty.

Conclusions:

1. a. Measurements of winds by rocket trails have shown that a monsoonal circulation exists in the 60-100 km region. Westerly winds in winter are replaced by easterly winds in the summer with the transitional period around the vernal equinox.

b. A region of maximum shear occurs near the 100 km level with shears in excess of 70 meters/sec/km. This exists in many localities and during different seasons.

c. A region of minimum wind. exists just above the level of maximum shear.

2. On the basis of data from meteorological rockets fired at Cape Canaveral, the following conclusions may be drawn:

a. Variability of the winds increases from 25 to 60 km. The mean wind speed differences range from approximately 2 mps at 25 km to 8 mps at 65 km.

b. For the summer period studied, the mean wind showed a maximum near the 50 km level.

c. A deep easterly flow exists over Cape Canaveral up to 50 km, with little change of direction and small meridional components.

d. Winds veer sharply above 50 km, becoming westerly above 60 km. Direction variability is large above 60 km because of this veering of the wind with height.

References:

1. Manring, E. J., Bedinger, Pettit, & Moore "Some Wind Determinations in the Upper Atmosphere Using Artificially Generated Sodium Clouds". J. Geophysical Research 64 587-591.

2. Ainsworth, J; Fox, & LaGow, "Measurement of Upper Atmosphere Structure by Means of Pitot-static Tube". NASA, Wash., D. C. Tech Note D-670.

3. Blamont, J., "Nuages Artificiel de Sodium et Vitesse du vent." Proc, first Int'n'l Space Science Symposium, Nice, 11-16 January 1960.

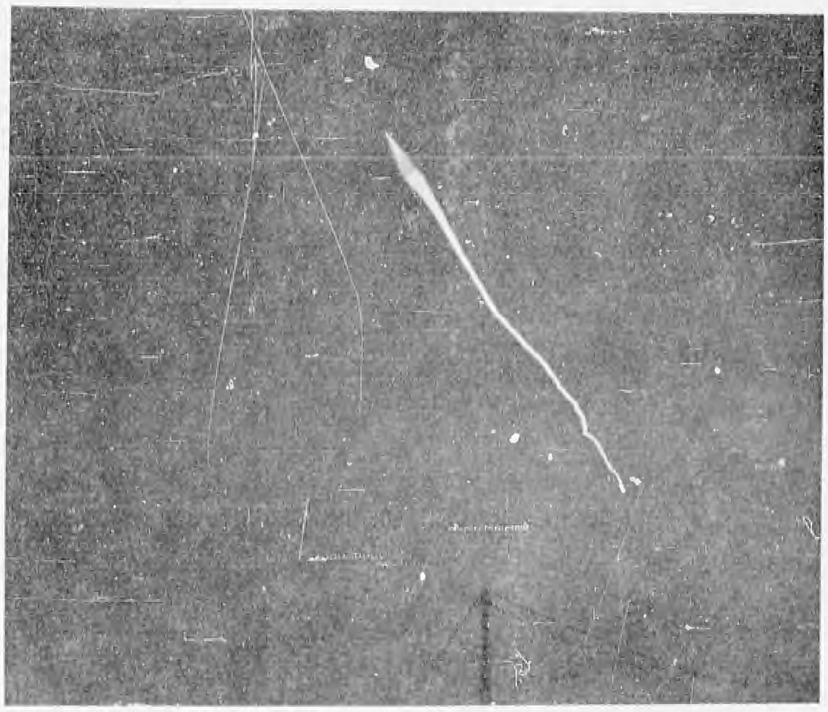


Fig. 1

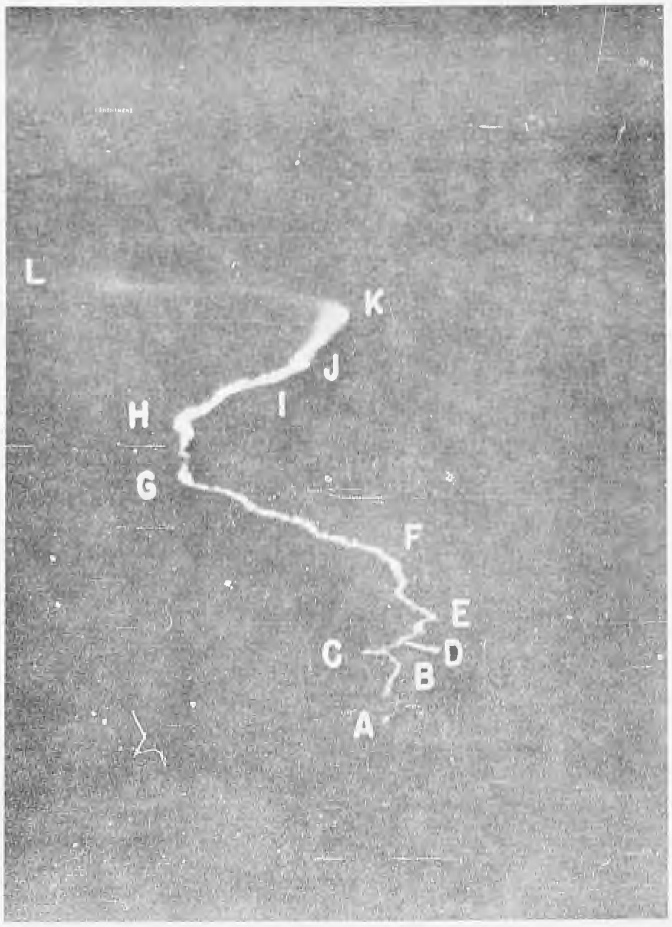


Fig. 2

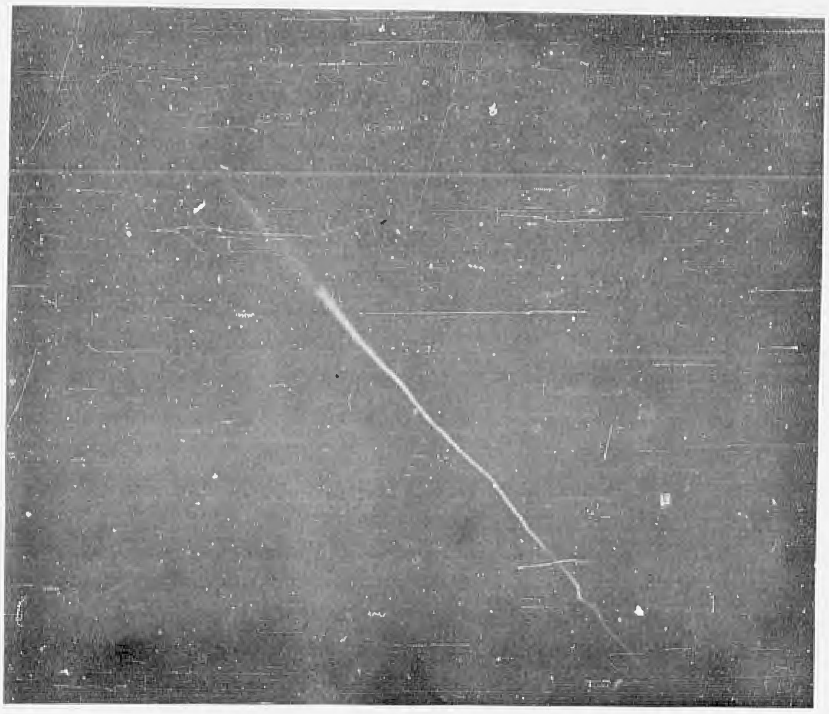


Fig. 3

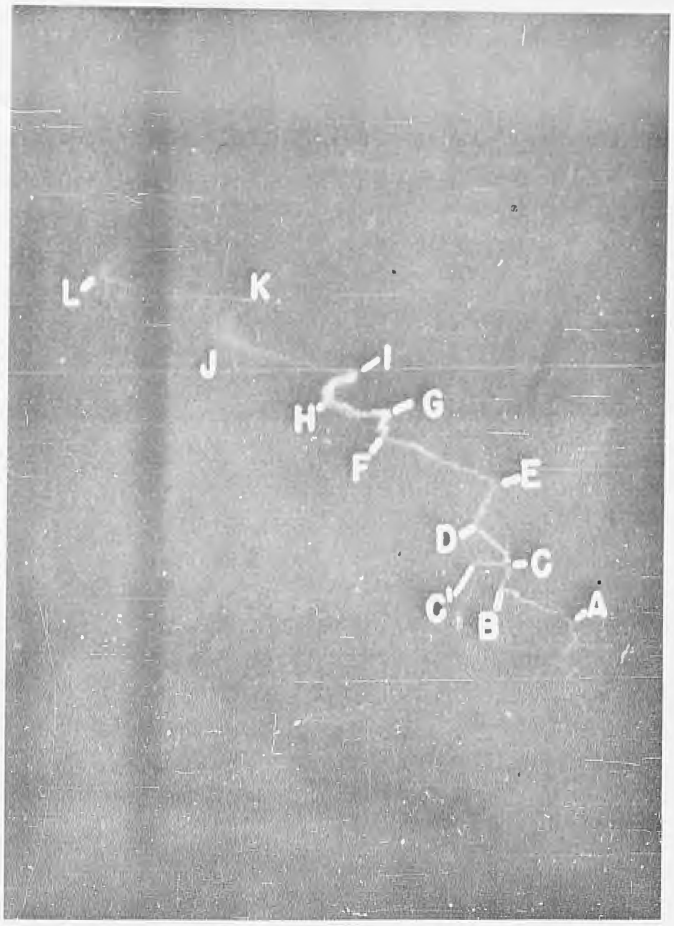


Fig. 4

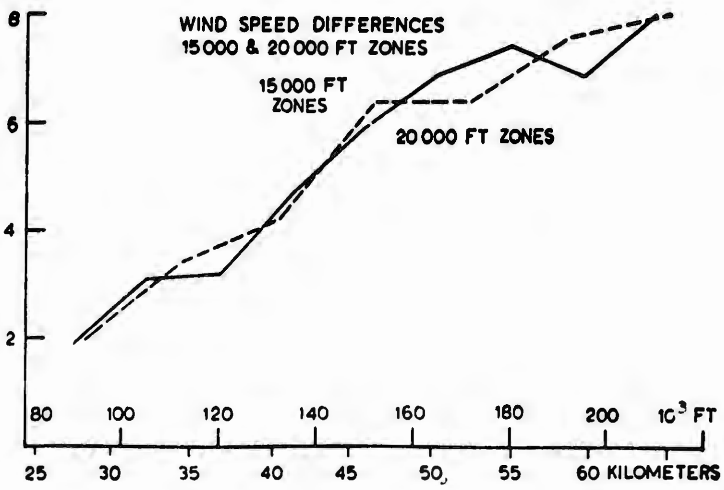

WIND SPEED DIFFERENCES
15000 & 20000 FT ZONES
15000 FT ZONES
20000 FT ZONES
8
6
4
2
80 100 120 140 160 180 200 10³ FT
25 30 35 40 45 50 55 60 KILOMETERS

Fig. 5.

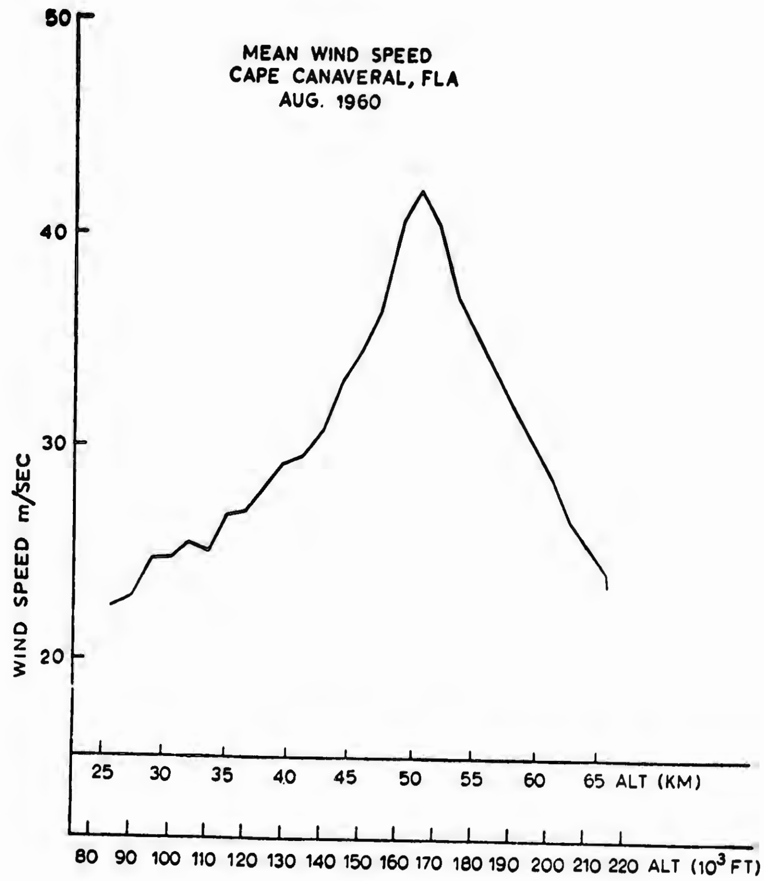
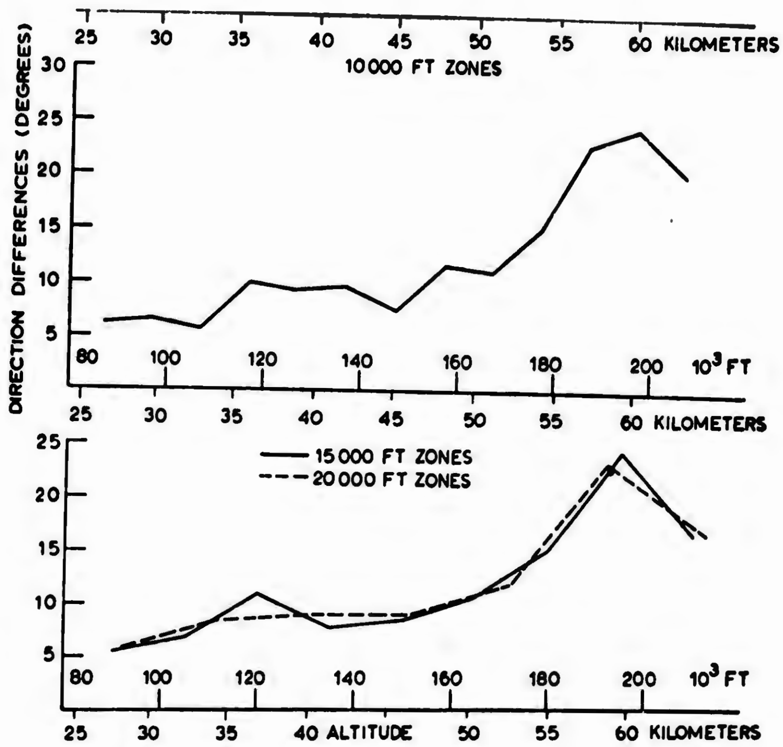

MEAN WIND SPEED
CAPE CANAVERAL, FLA
AUG. 1960
50
40
30
20
WIND SPEED m/SEC
25 30 35 40 45 50 55 60 65 ALT (KM)
80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 ALT (10³ FT)

Fig. 6.



WIND DIRECTION DIFFERENCES Fig. 7

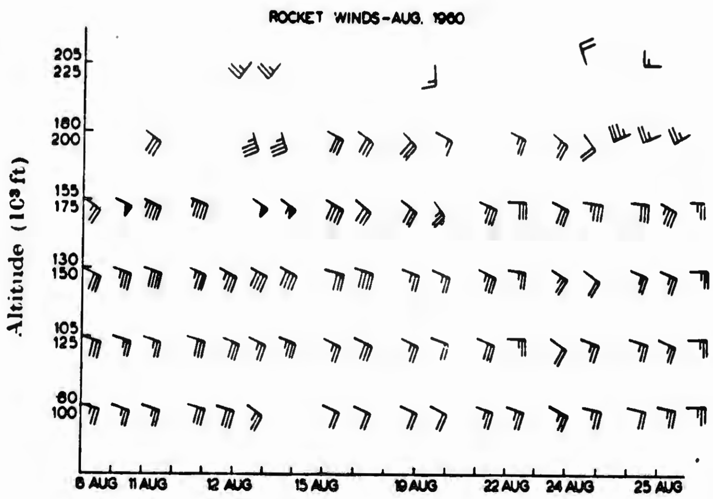


Fig. 8.

A BASIS FOR THE LONG RANGE PLANNING OF ARMY RESEARCH

G. H. McCLURG
RESEARCH PLANNING DIVISION
ARMY RESEARCH OFFICE
WASHINGTON, D. C.

I. PURPOSE.

The investigation upon which this Paper is based was undertaken in order to determine in what ways, and to what extent, the overall planning of Army research can be a meaningful and useful activity; to produce a rationale or guide for research planning; and to indicate how such planning may be oriented toward the long-range interests of the Army. This Paper is intended to present a discussion of some results of this investigation, to date, principally on the nature of research planning in general and on the possibility of relating it to operational goals.

II. WORD MEANINGS.

Dictionary definitions of the words "plan", "program" and "schedule" are not helpful, even for a qualitative inquiry. For the purposes of this discussion, the following meanings will be adopted:

PLAN: A guide to decisions on actions which are to produce desired results in the future. A Plan will include guidance for the allocation of resources to needs.

PROGRAM: A specific allocation of resources to needs, such as a research program for a particular Fiscal Year. A Program is derived from an R&D Plan by the application of particular input data and the exercise of judgment in evaluating criteria and carrying out the decision process.

SCHEDULE: A specific allocation of resources toward meeting a particular requirement over a period of time, such as a schedule of expenditures for a research project from year to year.

III. SCOPE.

Within the area of Army Research, several kinds of planning can occur, depending on what is being planned. This discussion will be limited to planning the scientific and technical content of research programs. It will not be concerned with fiscal planning aimed at obtaining the money resources needed to implement a proposed research program, or with financial management planning aimed at the orderly commitment and expenditure of money on research projects, so as to best conserve and utilize these resources within the applicable laws and regulations. All planning considerations not requiring scientific assessment of program content will be considered here as potential limitations or boundary conditions affecting the freedom of action of the scientific planner. Only basic and applied research are being considered, exclusive of supporting research for specific development contracts.

IV. PLANNING AS A PROCESS.

Since the determination of the scientific content of a research program is obviously a "process" of some sort, we might approach research planning by considering the plan for a process in general:

A. Assessment of resources.

B. Formulation of needs.

C. Comparison of resources with needs, using a language expressing the basic goals of the organization for which the plan is being devised, so that its problems (needs) can be matched to potential solutions selected from the set of resources.

D. If resources are insufficient, the Plan should provide guidance for improving resources or the methods of applying them, or for decreasing or simplifying needs, so as to obtain a balance.

E. If resources are still insufficient, then theories of value and cost must be applied so that a "best" allocation of available resources to the most important needs can be made.

F. Any final inability to meet minimum needs will be reported to a higher organizational level, where more resources may be available for allocation or where more alternatives may be open in reducing or modifying needs.

Even though this outline seems much too definitive to apply to research, scientific administrators must be using some such form of decision process, and it may be worthwhile to make comparisons.

V. THE PLANNING OF RESEARCH.

A. Resources.

At the input to the research process, resources are in the form of information, in the scientific literature and in the minds of research scientists. Output is also information, but higher in quality or quantity, better organized, or more useful. Eventually, this information becomes a part of the "state-of-the-art". Diffusing from person to person and into the permanent records of science, it forms the basis for new or improved technologies which enable the Army to improve its abilities. The assessment of research resources, therefore, does not seem to be a conceptual difficulty, although it faces the major problem of how to classify, abstract, store, retrieve and utilize scientific information, as well as the problem of how to efficiently utilize the capabilities of creative scientists through existing bureaucratic organizations.

B. Needs.

The formulation of needs or goals against which to consider or assess the relative value of proposed research efforts is a more difficult problem. Army research is undertaken with the intent of increasing the Army's capabilities, but existing statements of operational and equipment objectives seem inappropriate as goals for research planning. Some research guidance can be obtained by analyzing the existing requirements for new weapons systems to see where additional applied research may be needed, but little guidance can be expected from this source on how to assess the military implications of the more basic kinds of research which will lead to weapons system concepts of the future.

C. Comparison of Resources and Needs.

At present, there is a strong and increasing interaction between the rapid evolution of methods of warfare and the exponential advances of science and technology. This suggests a dual approach to Army research planning, in which each significant scientific advance would be related to all the basic Army operational capabilities which it might improve and each needed operational capability would be related to all the resources of science, both present and predicted. However, there is considerable difficulty in making these associations directly between the language of science and that of officially-approved statements of Army operational requirements. The latter need to be reasonably specific in order to form the basis for specific actions, while fundamental scientific discoveries often have broad military implications. It

may be desirable for Army research planners and technological forecasters to key their activities to an informal language in which a more fundamental list of Army operational capabilities can be compiled. It may be possible to construct such a list to represent all the basic operational needs of an Army, regardless of changes in warfare methods, world conditions or the status of science.

D. Army Operational Capabilities.

In determining what kind of entities might make up such a list, several characteristics seem to be important:

1. Long Useful Life.

a. Avoidance of military operational terms which imply specific tactical or organizational concepts, since these may change rapidly.

b. Clarity and definitiveness sufficient to resist the establishment of artificial similarities between distinct operational functions.

2. Operational Purity.

Terms descriptive of pure function or operation, with no words such as "radio communication" or "vehicular mobility", which imply a particular technological method of accomplishing the function.

3. Logical Completeness.

Entire outline to include all the basic operational needs of an Army. List to be reasonably expandible, in whole or in part, to attain greater detail, without illogical overlapping.

4. Semantic Effectiveness.

Universality of interpretation or impact of the words used on the people who can be expected to use them, regardless of organizational level.

5. Heirarchical Logic.

All logical alternative methods of attaining a particular operational capability to be equally obvious at each point of the outline.

6. Simplicity.

List to comprise no more than 3 or 4 heirarchical levels, and to use a minimum of coined words or common terms having

special meanings.

E. Use of Operational Capabilities List.

An adequate list of this kind could be used to classify both (1) the existing statements of Army requirements and (2) technological forecasts and proposed research programs, thus bringing the Army's research needs and potential resources together for consideration. If a considerable scientific potential were to appear in support of an operational capability for which there was no formally-stated need, the planner could investigate to see if an unrecognized requirement existed; or, if a stated need seemed to have no potential scientific solution, a special search could be made for one, or for some alternative.

F. Example of Operational Capabilities List

One of several ways of constructing such an Army Operational Capabilities List is illustrated in Fig. I. This is based on the assumption of 3 major capabilities needed by an Army: Force-producing elements, the information and communications needed to use them intelligently, and the ability to control environment and maintain support for continued operations. The element of Force consists of capabilities for physical destruction, the ability to apply threats based on Force, and the ability to neutralize enemy force elements by essentially non-violent means. The factor of Responsiveness represents the ability to obtain and utilize information, to make combat decisions accordingly and to communicate combat information needed in Army operations. The Endurance factor is involved with providing the conditions for human effectiveness in combat, with the utilization of passive defense, and with the provision of general support needed for continuing operations: power sources, military construction, logistics, general purpose information systems and the like.

Other approaches to building such a capabilities list are possible, a number of these have been tried, and more ideas along this line from Army scientists interested in research planning will be welcome. The principal requirements are to make the set of categories complete and easily separable, so that all Army operational capabilities will find a logical place in the list; to use words which are purely operational or functional in nature; and to keep the list reasonably short and simple, so it will be easy to use.

G. Values, Costs and Criteria.

1. Of the recent publications on military logistics and development planning, one of the most useful is Charles Hitch's book "The Economics of Defense in the Nuclear Age"(1) His discussions of the criterion problem and of incommensurables and uncertainty point out the dangers of an oversimplified approach to the

estimation of values and costs. In planning research, these concepts are more subjective than for development, but they should be considered regardless of the uncertainty. Useful criteria of value might include:

a. Degree to which the research is related to an operational requirement, its probable effectiveness in helping to meet the requirement, and the relative importance of the requirement with respect to others.

b. Proven ability of the investigator, relative importance of the work within its scientific area.

c. Nature of the work, whether it represents a minor extension of a well-known field or a pioneering effort in a new and relatively unknown area.

d. Breadth of applicability: whether results are likely to affect one discipline only or to have broad implications in many scientific fields.

2. For research, scales of cost can probably be merged with those of value, with the cost of a research effort represented by the (negative) value of giving up an alternative use for the resources involved. It may be useful to first divide the field intuitively into three parts:

a. High pay-off or promise, low cost.

b. High promise, high cost, or low promise, low cost.

c. Low promise, high cost.

Items in category a. can then be included, those in category c. excluded, and the decision process applied only to group b. If there is a budget ceiling and if the values of individual items affect each other, it may be necessary to reconsider the value of the selected items as a group, after the allocation has been made. The absolute as well as the relative values and costs may be significant and require consideration as upper or lower limits to the decision process. Probable rate or acceleration of progress may be a factor in new or exploratory areas.

H. Organizations, People and Motives.

Research characteristically involves a greater ratio of individual effort to team effort than development. Research scientists, as a group, tend to be individualistic, personally creative and deeply ego-involved in their work, and their effectiveness is likely to be adversely affected by the frustrations

encountered in a bureaucratic environment. Since the effectiveness of research is sensitive to organizational factors, the research planner will need to seriously consider the possibility of improving organization for research in order to upgrade the value of research resources. The question is worth a more extended discussion than can be given here, and References (3) (4) and (5) are recommended. There have been two main trends in Organization Theory, a traditional approach based on organization as a method of controlling and directing behavior, and the more modern view of organization as a method of effectively coupling the abilities of individuals to the achievement of organizational goals. The latter seems better adapted to research management, but the bureaucratic structure which became firmly established during the early days of the first Industrial Revolution, the rise of specialization and of mass production, still presents a major difficulty between the creative scientist and the scientific administrator. Perhaps the best answer is in a dual system: the scientists as individuals or small-group workers to provide the creative personal contributions and the scientific administrators as a coordinated group to provide the conditions necessary for creative work and to arrange to utilize the scientific output for the Army's benefit.

VI. CONCLUSIONS.

A. The overall planning of Army research can be a meaningful and useful procedure, and normal planning methods can be adapted to the purpose.

B. For use in research planning, a permanent list of Army operational capabilities can probably be constructed, which will provide a more direct and fruitful link between the resources of science and the long-range interests of the Army.

C. Technological forecasting and research planning are new and important disciplines requiring the development of a specialized professional competence within the Army.

References

(1) Hitch, Charles J. and McKean, Roland N., "The Economies of Defense in the Nuclear Age", RAND Corporation Report R-346, March 1960.

(2) Klein, Burton, "A Radical Proposal for R&D", FORTUNE, May 1958, pp. 112 ff.

(3) Haire, Mason, (Ed.), "Modern Organization Theory", Wiley, New York, 1959.

(4) Argyris, Chris, "Understanding Organizational Behavior", Dorsey Press, Homewood, Ill., 1960.

(5) Merton, R. K., et al, (Eds.), "Reader in Bureaucracy", Free Press, Glencoe, Ill., 1960.

FIG. I

ARMY OPERATIONAL CAPABILITIES

PRINCIPAL CAPABILITIES — MAJOR FUNCTIONS — ELEMENTS — COMPONENTS

- **I FORCE**
  - PHYSICALLY DESTROY OR DAMAGE
    - WEAPON SYSTEMS
      - Damage-Producing Devices
      - Delivery Means
      - Integral Support
    - TARGET SURVEILLANCE
      - Search
      - Detection
      - Identification
      - Location
      - Tracking
      - Fire Control
      - Damage Assessment
  - APPLY THREATS BASED ON FORCE
    - Intimidation
    - Harassment
    - Sabotage
    - Infiltration
    - Guerrilla Warfare etc.
  - NEUTRALIZE OPPOSING CAPABILITY
    - Nonlethal Agents
    - Psychological Warfare
    - Persuasion
    - Propaganda
    - Subversion
    - Military Occupation, Control & Gov't of Land Areas
    - Electronic Warfare, Jamming, Deception
    - Economic, Political, Sociological Actions with Military Implications or Intentions

PRINCIPAL CAPABILITIES — MAJOR FUNCTIONS — ELEMENTS

- **II RESPONSIVENESS**
  - OBTAIN INTELLIGENCE
    - Information on Past, Present, Future Enemy Capabilities, in Advance of Hostilities
  - CONDUCT COMBAT SURVEILLANCE
    - Locate, Identify Enemy Ground Forces
    - Obtain & Use Current Information
    - Make Command Decisions
  - PROVIDE COMMAND COMMUNICATIONS
    - Convey Command Decisions to Combat Units
    - Feedback Information to Commanders
    - Provide Essential Intercomm between Military Combat Units for Command Purposes
- **III ENDURANCE**
  - ESTABLISH HUMAN EFFECTIVENESS
    - Clothing, Food, Personal Equipment
    - Environmental Protection
    - Man/Machine Compatibility
    - Training
    - Selection, Assignment, Utilization
    - Medical, Psychological
    - Morale, Leadership
    - Organization, Tactics, Strategy
    - Personal Protection vs. Enemy Force
  - UTILIZE PASSIVE DEFENSE
    - Collective Personnel Protection
    - Camouflage, Deception
    - Defensive Construction
    - Armor
    - Warning Systems
  - PROVIDE FOR COMBAT SUPPORT
    - Logistics: Storage, Transportation, Maintenance
    - Information Systems, Data Processing, General Purpose Communications
    - Military Construction: Roads, Bridges, Installations, etc.
    - Power Sources

EFFECT OF MAFENIDE (SULFAMYLON[R]) AND PENICILLIN IN TOPICAL THERAPY OF C. PERFRINGENS-CONTAMINATED EXPERIMENTAL WOUNDS*

JANICE A. MENDELSON, MAJOR, MC
U. S. ARMY CHEMICAL RESEARCH & DEVELOPMENT LABS.
ARMY CHEMICAL CENTER, MARYLAND

I. INTRODUCTION

In military surgery the principle has arisen that missile wounds involving large soft tissue masses (especially muscle) require surgery. This applies particularly to wounds of the buttocks, thigh, and calf (Figure 1). These wounds can be fatal, even though no "vital" structure has been damaged (10).

The operation consists of opening the skin and fascia immediately and removing all muscle tissue that appears non-viable (Figure 2). The resulting wound is then left open for several days and then usually requires another operation to close it. The chief reason for the extensive surgery is to prevent the development of infection. This refers particularly to the anaerobic infections, especially gas gangrene. These organisms grow best where the oxygen supply is diminished, and produce their toxic manifestations particularly in the presence of dirt, debris, and dead muscle tissue (1, 2, 21). The longer the delay between wounding and surgery, the higher the incidence of gas gangrene (10). The exact mechanism whereby the development of gas gangrene in an extremity results in the death of the patient is not known; it is quite certain, however, that the infected wound is the source of the toxicity (3, 12, 13, 14, 16, 17, 23, 24).

Present military techniques may result in less prompt surgical care than was possible in World War II and the Korean conflict. The use of radioactive materials may require injured people to remain under cover for 24 hours or longer. The emphasis on use of small independent airborne units, such as special forces, who may be situated behind enemy lines, renders prompt adequate surgery unfeasible in many instances. The results of adequate surgery may temporarily incapacitate the patient more than the missile wound itself. Saturation of medical facilities in a mass casualty situation also might result in delayed or inadequate surgery.

---

* Experimental animals used in this study were handled in accordance with the principle of laboratory animal care as promulgated by the National Society for Medical Research (reference AR 70-18 20 Nov 61).

Therefore, studies have been done in our laboratory to see if there is a simple way of prolonging the period of time during which surgery will be effective, and preventing deaths from gas gangrene when prompt adequate surgery cannot be done.

Because of the conditions under which they would be employed, it was necessary to consider treatment methods requiring a minimum of skill and equipment. We therefore considered topical therapy. This had been attempted during World War II and at the end of the war had been abandoned as of little value. Sulfonamides and, toward the end of the war, relatively small doses of penicillin were used. When the reports evaluating their value are carefully reviewed, one notices that data concerning several types of sulfonamides, applied largely by dumping powder or crystals into the wounds, were often included together (4, 18). One of the most frequently used topical sulfonamides was the relatively ineffective sulfanilamide (15). Many of the sulfonamides caked in the wounds, acting as foreign bodies, and interfering with drainage. In some instances no attempt to evaluate the extent of underlying damage was done, and the sulfonamide essentially merely placed on the surface of an entrance wound (7, 20). However, several studies indicated beneficial results from topical therapy (4, 6, 7, 8, 11). All of the sulfonamides in common use were inhibited by p-amino benzoic acid which is found in infected wounds, so the drugs soon lost their effectiveness (22).

During World War II, the Germans used another sulfonamide, which they called Marfanil, for the prevention of wound infection. Some of their reports indicated that it was effective against the organisms of gas gangrene. Unlike the other sulfonamides, Marfanil is not inhibited by PABA (9, 22). By the time Marfanil became available to the American and British medical services, penicillin was just emerging as a new "wonder durg" and relatively little attention was given to Marfanil. Some laboratory tests were done, with conflicting results.

For about eight years in our laboratory, studies have been done using a standardized massive wound of the anaesthetized goat (5, 12, 13, 14, 16, 17, 19, 23). This wound, produced by a high explosive charge, simulates by actual test that due to a large high velocity missile. There is an open comminuted fracture of the femur, extensive muscle damage, and the femoral artery is severed (Figure 3). When untreated, all of these goats spontaneously develop a fatal infection with gas gangrene organisms (C. perfringens predominantly) (12, 13, 14). Amputation is required to save their lives. It has been found possible to delay the time of amputation by 24 hours if the goat is treated with large doses of penicillin, parenterally or topically. However, penicillin alone will not permit recovery of the animal (5, 16, 17, 19).

Antibiotics are not a panacea. Many of them result in the development of antibiotic resistant bacteria; some produce severe allergic reactions. If nothing better is available, their use is of

course justifiable. For these reasons, it seemed worthwhile to reconsider the use of chemotherapeutic agents in addition to the antibiotics. Therefore Marfanil was included in these studies. In the United States its trade name is Sulfamylon; its generic name mafenide. Chemically it is 4-(aminomethyl) benzenesulfonamide hydrochloride.

II. PROCEDURE

A total of 149 goats was used in this study. Six treatment groups were compared. These were:

1. Untreated controls
2. Penicillin spray
3. 5% mafenide†
4. 20% mafenide
5. 5% mafenide + penicillin
6. 20% mafenide + penicillin

Each treatment was tested in three to five subgroups of six goats each, with simultaneous controls of six goats each. This was to permit evaluation of minor potential variables, and because some of these were part of other studies not being reported at this time. As will be discussed, with one exception the subgroups proved statistically similar and the results could be combined into one treatment group each. There were five groups of untreated controls, each containing six goats. There were five groups of six goats each treated with penicillin spray, 2 million units (1.2 gm) in 50 ml sterile distilled water per treatment. The treatment was applied immediately after wounding and twice daily thereafter as long as the animal survived.* All treatment was done by spraying the treatment solution in a DeVillbis atomizer using nitrogen under ten pounds of pressure as a propellant for uniform force and rate. Three groups of six animals each were treated with 5% mafenide spray applied in the same way as the penicillin. Five groups of six goats each were treated with 20% mafenide spray. With one exception, the treatment was applied as in the preceeding groups. The exception was that six animals were treated after a five hour delay instead of immediately following wounding, the subsequent treatment being as in the other groups.

Both of the other two large treatment groups had three components of six goats each. In one of these, 5% mafenide was mixed with 2 million units of penicillin per treatment, and in the other 20% mafenide and 2 million units of penicillin was applied per treatment.

III. RESULTS

The variables among the subgroups receiving each treat-

† Mafenide in the form of Sulfamylon(R) was supplied for these studies by Winthrop Laboratories, New York

* Treatment was usually omitted over weekends

ment included such factors as the number of times the animals were disturbed for laboratory tests, type of diet, and the time after wounding that they were removed from the treatment-observation carts and placed in the goat pen. However, statistical analysis revealed that, with one exception, there were no significant differences in results among the subgroups and they could therefore be combined into the major treatment groups for analysis. The results of the subgroups are shown in Figure 5. For brevity and ease of comparison the combined results are illustrated in Figure 6. The exception to the validity of combining the subgroups is the 5% mafenide group, which contained three subgroups. The first group of animals consisted of 12 goats with a median survival time of about 133 hours. Half of these goats lived over 400 hours. The other groups of animals treated with 5% mafenide both had median survival times of about 70 hours, with no animal living over 400 hours. In the latter two groups the mean and median survival times were at least twice that of the untreated controls, approaching results obtained with penicillin spray in the same studies. It is not yet known whether the former or the latter two groups are more representative of the 5% mafenide effect. In Figure 6 this is illustrated by the wide spread of survival times in the 5% mafenide group. The six longest survivors were all members of the first group studied.

The most dramatic effect was seen with the use of 20% mafenide plus penicillin. The first of these animals to die outlived all of the controls and all of the goats treated with penicillin alone. None of these animals died before 150 hours after wounding, and 15 of the total of 18 lived over 200 hours. This is remarkable because in all of the previous related studies done in this laboratory, the only goats (except those treated surgically) which survived over 100 hours were three or four antibiotic treated animals who accidentally acquired superimposed maggot infestation (5, 12, 13, 14, 16, 17, 19, 23). Each of the groups of goats treated with mafenide alone or in combination immediately after wounding and twice daily thereafter had at least one animal living more than 900 hours after injury, there being a total of 10 goats living at least this long. In contrast, although in this study the untreated controls tended to live longer than in most of the previously reported studies from this laboratory, none lived over 51 hours, the mean and median survival times of the combined group being about 34 hours.

Suppression of bacterial growth, as evaluated by quantitative cultures for *C. perfringens* and *Staph. aureus*, appeared to be almost complete in the wounds treated with the mafenide-penicillin combination, and, in some of the subgroups, in wounds treated with mafenide alone. As has been determined before, penicillin used alone suppresses bacterial growth initially, but by 16-48 hours, positive cultures may be found. The survival time of these animals has been previously found to bear a mathematical relationship to the dosage of *C. perfringens* to which they are exposed. *Staph. aureus* is measured

merely for comparison. There are also other organisms present in these wounds, but they do not contribute quantitatively to the survival time of the animals. Cultures reported as "0" represent less than 1000 organisms per ml of wound exudate. It is not inferred that these wounds are actually sterile. Two of the very late survivors, whose wounds had almost healed, had apparently superficial staphylococcal wound infections until the wound healed. The depths of the wounds in the long surviving animals cannot be adequately sampled, because the wound forms a massive dry crust which is not otherwise disturbed until it separates spontaneously. Further studies involving sacrifice of long-term survivors are being done at present to permit evaluation of the extent, if any, of non-lethal bacterial growth in the tissues deep to the wound surface. These are not completed at this time.

There were three types of wound responses. The wounds of untreated animals remained moist and the muscle soft and dull pink in color (Figure 4). If the goat lived long enough, by about 30 hours the muscle tissue became slimy and grey and a foul brownish fluid began to be present in the wound. The wounds of the penicillin treated animals also remained moist. By about 36 hours after wounding, the muscle tissue looked pinker and cleaner and firmer than that seen in the untreated controls. Eventually, however, the brownish wound exudate developed and the muscle tissue became yellowish-grey in appearance (Figure 7). In contrast, the wounds treated with mafenide alone or mafenide with penicillin appeared very different. By about 36 to 72 hours after wounding, the entire wound area was dry and greyish-black (Figure 8). By about 45-65 days after wounding the entire wound separated off as a massive crust, leaving healthy granulation tissue beneath (Figure 9). In some instances the crust separated gradually, with the same end result. These animals were not restrained or splinted in any way after being placed in the goat pen. It was found that transfer to the goat pen about 25 hours after wounding was the optimum time for long survival, this being one of the variables tested in the subgroups. Goats, like other ruminants, develop serious effects from immobilization and for proper nutrition have to be able to assume both the standing and the recumbent positions. Therefore the severely fractured femur received no therapy. However, in some instances, although the orthopedic results were far from optimum (Figure 10), the bony continuity was restored (Figure 11), and a weight-bearing functional extremity resulted (Figure 12). In others of the "recovered" animals, as would be expected, non-union of the fracture resulted, although the soft tissue wound healed. In two of the animals, the residual blood supply of the injured extremity was so poor that the limb developed "dry gangrene" (without evidence of infection) and spontaneously separated, after which the soft tissue wound healed uneventfully. The long term survivors are lively and apparently healthy. In those with survival of the leg, there is usually at least stiffness of the hip joint and some shortening of the injured extremity, but all of the goats eat well and run about freely.

IV. DISCUSSION

The complete recovery of some of the animals with no surgical therapy whatsoever was entirely unexpected. It appears that the medication prevented the multiplication of C. perfringens in the wounds, and prevents the absorption of toxic substances from the wound which are associated with C. perfringens infection. It is probable that the presence of actively growing C. perfringens is essential for the uniformly fatal course of the untreated wound.

The present study does not evaluate the treatment of an established gas gangrene infection. The group of six goats with initial treatment five hours after wounding had mean and median survival times only slightly better than the untreated controls, but the results of treatment of this small a group may not be conclusive. Further testing is in progress to evaluate the efficacy of delayed treatment more adequately. This study does show, however, that the immediate and repeated use of topical mafenide as the sole therapy can permit recovery from an otherwise fatal massive wound, and that this parallels prevention of the development of recoverable C. perfringens in the wounds.

V. SUMMARY AND CONCLUSIONS

Topical mafenide as the sole therapy can permit recovery from an otherwise fatal experimental massive open wound. The efficacy of treatment parallels the prevention of development of recoverable C. perfringens in the wound, this organism otherwise being always present in the wound by 9 hours after wounding. Topical penicillin alone prolongs survival time but does not permit recovery. The addition of penicillin to mafenide appears to increase the efficacy of the treatment. The applicability of these results to human wounds remains to be studied.

VI. LITERATURE CITATIONS

1. Altemeier, W.A.: The Bacteriology of War Wounds, Surg. Gyn. & Obst. 75: 518-533, 1942.
2. Altemeier, W.A. and Furste, W.L.: Gas Gangrene: Collective Review, International Abstracts of Surgery (Surg. Gynec. & Obst.) 84: 507-521, June, 1947.
3. Balch, H.H. and Cranley, O.H.: Observations on the Pathogenesis of Clostridium Welchii Myonecrosis, Ann. Surg. 146: 86-97, 1957.
4. Bentley, F.H. and Thomson, Scott: Control of Infection in Recent Wounds by Surgery and Local Chemotherapy, Brit. M.J. 1: 471-474, April 7, 1945.
5. Brinkley, F.B., Schneider, R.F., Stevens, A.M. and Lindsey, D.: Effect of Local Penicillin Spray on Survival Time Following a Massive Open Wound, Antibiot. Ann., 526-530, 1959-1960.
6. Brown, J.J.M.: Early Closure of Soft Tissue Wounds with Chemotherapeutic Agents: A Comparative Study of Sulphanilamide & Penicillin, Brit. J. Surg. 32: 140-143, 1944.
7. Crile, George, Jr.: Experiences of the Surgical Service of the U. S. Naval Hospital, Auckland, New Zealand, with Casualties from Initial Solomon Islands Engagement, U. S. Nav. Med. Bull. 41: 306-324, March, 1943.
8. Ferguson, D.K.: Observations on the Treatment of Battle Wounds Aboard a Hospital Ship, U. S. Nav. Med. Bull. 41: 299-305, March, 1943.
9. Hamre, D.M., Walker, H.A., Dunham, W.B., et al: Chemotherapy of Infections with Cl. Perfringens in Mice by Homosulfonamides, Proc. Soc. Exp. Biol. & Med. 55: 170-173, March, 1944.
10. Holmes, R.H., Enos, W.F., Jr., and Beyer, J.C.: Medical Aspects of Body Armor Used in Korea, J.A.M.A. 155: 1477-1478, 1954.
11. Langley, F.H. and Winkelstein, L.B.: Gas Gangrene: A Study of 96 Cases Treated in an Evacuation Hospital, J.A.M.A. 128: 783-792, July, 1945.
12. Lindsey, D.: Gas Gangrene: Clinical Inferences from Experimental Data, Am. J. Surg. 97: 582-598, May, 1959.
13. Lindsey, D., Noyes, H., et al: The Role of Clostridia in Mortality Following an Experimental Wound in the Goat, Surgery 45: 602-622, April, 1959.
14. Lindsey, D., Wise, H.M., Jr., Knecht, A.T. and Noyes, H.E.: The Role of Clostridia in Mortality Following an Experimental Wound in the Goat: 1: Quantitative Bacteriology, Surgery 45: 602-611, April, 1959.
15. Lyons, C.: An Investigation of the Role of Chemotherapy in Wound Management in the Mediterranean Theater, Ann. Surg. 123: 902-924, 1946.
16. Mansberger, A.R., Jr., Ochsner, E.W.A., Jr., Jacob, S., Oppenheimer, J.H. and Gillette, R.W.: A New Preparation for the Study of Experimental Shock from Massive Wounds: II: Evaluation of Various Therapeutic Regimens with Special Reference to the Role of Antibiotics, Fluid Replacement, and Debridement, Surgery 43: 708-720, May, 1958.

17. Mansberger, A.R., Jr., Oppenheimer, J.H. and Gillette, R.W.: A New Preparation for the Study of Experimental Shock from Massive Wounds: III: The Effect of Local Wound Treatment on Survival Time, Surgery 43: 721-729, May, 1958.
18. Meleney, F.L.: A Statistical Analysis of a Study of the Prevention of Infection in Soft Part Wounds, Compound Fractures and Burns with Special Reference to the Sulfonamides, Surg. Gynec. & Obst. 80: 263-296, 1945.
19. Mendelson, J.A.: Topical Therapy as an Expedient Treatment of Massive Open Wounds, Surgery 48: 1035-1047, Dec., 1960.
20. Monroe, Clarence W.: Debridement - When & How Much? A Comparative Study of Battle Casualties, Bull. of U. S. Army Med. Dept. 77: 37-41, June, 1944.
21. Neel, H.B. and Cole, J.P.: Gas Gangrene in Amphibious Warfare in the Pacific Area, Am. J. Surg. 66: 290-299, Dec., 1944.
22. Northey, E.H.: The Sulfonamides and Allied Compounds, Reinhold Publishing Corporation, New York, 1948.
23. Noyes, H.E., Sherman, R., Wise, H.M., Jr., Knecht, A.T., Jr., and Lindsey, D.: The Role of Clostridia in Mortality Following an Experimental Wound in the Goat: II: Demonstration of Toxic Materials in Wound Exudate, Surgery 45: 612-616, April, 1959.
24. Stammers, F.A.R., et al: Section of Pathology: Discussion on the Toxaemia of Gas Gangrene, Proc. Royal Soc. of Med. 39: 291-296, 1946.

FIGURE 1. LEG AND THIGH WOUND (KOREAN CONFLICT)

FIGURE 2. DEBRIDED WOUNDS (SAME PATIENT AS FIG. 1)

FIGURE 3. GOAT THIGH IMMEDIATELY AFTER INJURY

FIGURE 4. UNTREATED WOUND 24 HOURS AFTER INJURY

SULFAMYLON

	TREATMENT	NUMBER OF GOATS	CULTURES (log 10): CL. perfringens (mean) 9 16 25 36 48 72 100 144	CULTURES (log 10): STAPH aureus (mean) 9 16 25 36 48 72 100 144	SURVIVAL TIME: MEDIAN	MEAN	RANGE	S D +	SURVIVORS: OVER 100 HRS NO.	%	OVER 200 HRS NO.	%	OVER 400 HRS NO.	%	OVER 600 HRS NO.	%
CONTROLS	Part I	6	3.3 4.6 6.9 7.3	4.0 5.4 4.1 4.6	35.66	37.62	32.33– 47.25	5.05	0	0	0	0	0	0	0	0
	II	6	2.2 5.4 7.2 7.2	3.5 5.6 6.2 7.3	34.32	36.29	26.42– 44.42	6.20	0	0	0	0	0	0	0	0
	III	12	2.1 3.0 5.2 6.8	2.8 4.0 5.6 5.0	23.87	30.22	3.33– 47.00	13.26	0	0	0	0	0	0	0	0
	IV	6	2.8 5.3 7.0 –	3.6 4.3 4.6	34.29	37.11	22.92– 51.08	10.50	0	0	0	0	0	0	0	0
	V	6	2.3 5.5 6.5	3.5 4.0 4.5	25.35	30.26	12.08– 43.50	3.79	0	0	0	0	0	0	0	0
	Total Mean	36	2.5 4.5 6.3 7.0	3.4 4.6 5.1 5.5	33.89	33.62	3.33– 51.08	9.80	0	0	0	0	0	0	0	0
PENICILLIN SPRAY	Part I	6	0.5 1.1 1.3 3.4 4.8	0 0 0 0 0	104.2	107.29	82.92– 137.50	19.78	4	67	0	0	0	0	0	0
	II	6	0 0.5 0.8 1.3	0.5 1.0 0 0	101.4	111.97	74 – 151.7	29.84	4	67	0	0	0	0	0	0
	III	(not tested)														
	IV	6	0 0 0 1.2 1.8	0 0 0 1.4 2.0	102.14	110.71	68.92– 136.87	27.36	4	67	0	0	0	0	0	0
	V	6	0 0 1.0 3.0 3.4 5.0	0 0 0 0.8 3.0 1.8	82.36	119.79	16.00–205.00	60.79	5	83	1	17	0	0	0	0
	Total Mean	24	0.1 0.6 0.7 2.4 3.0 2.6 5.0	0.1 0.2 0 0 0.7 2.5 1.8	101.30	112.44	16.00–208.00	37.83	17	71	1	4	0	0	0	0
5% SULFAMYLON	Part I, II	(not tested)														
	III	12	0 0.3 2.6 2.9 5.2	0 0.3 0.8 0.9 3.2	132.65 [o]	326.02 [x̄]	91.33–ALIVE		8	67	7	58	6	50	3	25
	IV	6	0 0 0 0 0 0	0 0 0 0 0 0	69.80	144.33	24.25–368.08	135.53	4	67	1	17	0	0	0	0
	V	6	0 0 0 1.2 2.2 2.0	0 0.6 0.8 2.6 2.8 1.3	73.12	121.97	11.00–216.33	79.78	3	50	2	33	0	0	0	0
	Total Mean	24	0 0.2 1.3 2.9 0.6 1.1 1.0	0 0.3 0.6 0.9 1.3 1.4 2.8 0	130.95 [o]	221.90 [x̄]	11.00–ALIVE		15	62	10	42	6	25	3	12
20% SULF	Part I	6	0 0 0 3.6 4.0	0 0 0.8 0.5 0	66.6	196.32	32.00–416.17	138.5	3	50	1	17	1	17	0	0
	II	6	0 1.5 3.2 0.6	0.5 0.5 3.1 2.6	64.90	96.28	36.67–152.17	46.41	2	33	0	0	0	0	0	0
	III															
	IV	6	0 0 0 0 0 0	0 0 0 0 0 0	154.90	157.95 [x̄]	36.25–ALIVE		4	67	3	50	1	17	1	17
	V	6	0 0 1.3 3.7 3.5 0	0 0 1.2 3.5 0 1.0	96.39	177.86	66.58–490.33	173.97	2	33	2	33	1	17	0	0
	Total Mean	24	0 0.4 1.1 2.1 2.6 1.8 0 0	0.1 0.1 1.3 1.6 1.2 0 1.0 0	108.28 [o]	144.07 [x̄]	36.67–ALIVE		11	46	6	25	3	12	1	4
20% SULF DELAYED	Part II	6	0 1.2 1.5 4.0	1.0 1.4 2.6 1.5	35.20	52.92	12.92– 88.08	35.46	0	0	0	0	0	0	0	0
5% SULF + PEN	Part I, II	(not tested)														
	III	12	0 0 0.4 0.9 1.3	0 0.6 0 0.7 1.5	355.36 [o]	403.44 [x̄]	17.40–ALIVE		10	83	8	67	5	42	3	25
	IV	6	0 0 0 0 0 0	0 0 0 0 0 0	115.06	234.56	20.08–319.42	201.94	4	67	4	67	2	33	0	0
	V	5	0 0 0 0.6 0 0	0 0 0.6 0.8 0 0	219.71 [o]	361.50 [x̄]	67.25–ALIVE		4	67	2	33	2	33	2	33
	Total Mean	23	0 0 0.2 0.9 0.3 0 0.9 0	0 0.3 0.1 0.7 0.4 0 1.0 0	206.47 [o]	350.38 [x̄]	17.40–ALIVE		18	75	14	58	9	39	5	22
20% SULF + PEN	Part I	(not tested)														
	II	6	0 0 0 0	0.5 0.5 0.5 0	341.4	457.5	156.33–841.00	292.61	6	100	4	67	3	50	2	33
	III	(not tested)														
	IV	6	0 0 0 0 0 0	0 0 0 0 0 0	484.80	499.28 [x̄]	345.75–ALIVE		6	100	6	100	6	100	3	50
	V	6	0 0 0 0 0 0	0 0 1.0 0 0 0.5	309.82 [o]	325.85 [x̄]	197.83–ALIVE		6	100	5	83	3	50	1	17
	Total Mean	18	0 0 0 0 0 0 0 0	0.2 0.2 0.5 0 0 0 0.2 0	376.28 [o]	425.10 [x̄]	156.33–ALIVE		18	100	15	83	12	67	6	33

x̄ mean survival of those animals that died
o median survival estimated from survival times of goats that died
+ S D - Standard Deviation

FIGURE 5. SUMMARY OF RESULTS

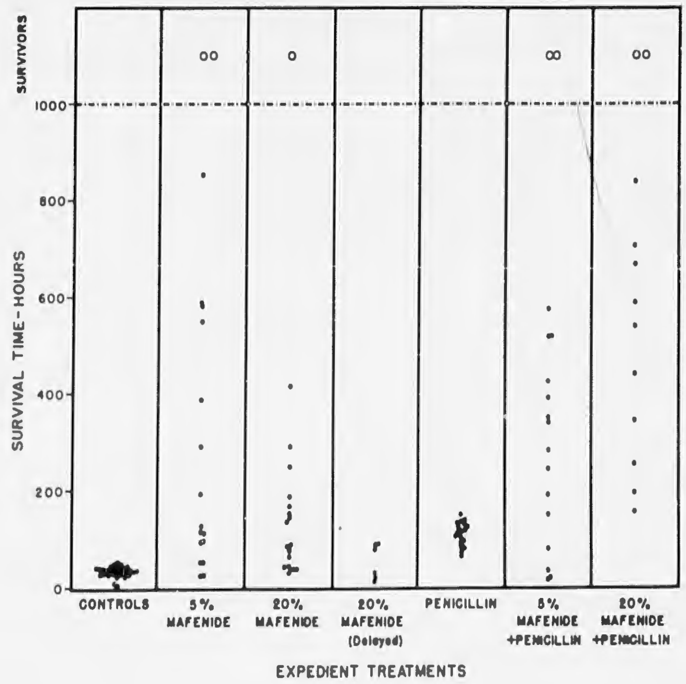


FIGURE 6. COMBINED SURVIVAL TIMES COMPARED

FIGURE 7. PENICILLIN TREATED WOUND (72 HOURS)

FIGURE 8. MAFENIDE TREATED WOUND (72 HOURS)

FIGURE 9. MAFENIDE TREATED WOUND - 65 DAYS

FIGURE 10. X-RAY 25 DAYS AFTER INJURY

FIGURE 11. X-RAY OF SAME GOAT AS FIG. 10 AT 133 DAYS

FIGURE 12. AMBULATORY GOAT, WOUND HEALING

LASER PROGRESS AND APPLICATIONS

HARRISON J. MERRILL
U. S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY
FORT MONMOUTH, NEW JERSEY

The LASER is a recent development resulting from studies and research in changes in the energy state of atoms. The word is coined from the initial letters which describe the phenomenon, "Light Amplification by Stimulated Emission of Radiation." The energy emitted from the LASER displays very unusual coherence characteristics never before encountered in optical phenomena. It has become an interesting phenomenon which has resulted in the publication of a considerable amount of scientific as well as popular literature.

The LASER is an outgrowth of basic research which produced a very low noise radio amplifier called the MASER. Extension of the MASER into the optical region was considered and postulated by Schwalow and Townes at Columbia University in an announcement published in December 1958.[1] This work at Columbia was initiated by the Signal Corps on a government contract and was later supported as a tri-service contract by the Army, Navy and Air Force.

By March 1961, exploratory research in the use of a LASER had progressed to the point where it was considered desirable that it be applied for military use. Projections in which the growth of the development of the LASER were forecast resulted in the finding that the LASER had considerable potential for range measurement in connection with surveillance equipment; and it was decided that a program be instituted to accomplish this.

In addition to the range measurement, there are a number of other potential applications, such as: in communications; transmission of power; illumination and guidance and control. However, these applications require developments which cannot be forecast, thus requiring further research. In the civilian economy, the LASER may be applied to replace eye surgery in order to correct detached retinas; in micro welding; drilling and machining and as a unique

light source for other scientific investigations. Although there have been many materials described which are suitable for LASER applications, the use of the ruby has unique characteristics which, up to the present time, make it outstanding as a source.

LASER OPERATION

Essentially, the LASER output is derived from a cylindrical cavity which is terminated at one end by a highly reflective surface and at the other end by a partially transmitting surface. The cavity may be either partially or completely filled by an active material which must have the characteristic that it fluoresces when excited. The active material may be a suitable crystal, such as aluminum oxide (ruby) doped with chromium, calcium tungstate, doped with neodymium, or calcium fluoride doped with uranium; a solid solution, such as glass doped with neodymium; or a gas, such as a neon-helium mixture. The action of the LASER, will be discussed specifically using the ruby as the active structure although other materials could have been used. The ruby is illustrated in Fig. 1 and is ordinarily about ¼" by 2" long. This ruby is not the gem type for it has less chromium ($\approx$ .05) doped in the crystal. Instead of the deep red we ordinarily associate with the ruby it is a light pink.

To get LASER action, we place the pink ruby in the cavity and illuminate it with a short and very intense pulse of light from a gaseous discharge tube. This tube is similar to electronic flash units commonly used by photographers. The energy transitions are shown in the lower left of Fig. 1. The horizontal lines represent energy levels within the ruby structure. The base line indicates the ground energy level, n_o. The transition between the next higher energy level, n_1, and n_o corresponds to the characteristic fluorescent red color of the ruby and the top line, n_2, indicates that corresponding to the chromium impurity absorption. Actually the line structure is divided into double levels, which are not shown in order to simplify this discussion. The energy states normally predominate at the ground level but, when the flash pumps light into the ruby, the energy states are excited by absorption of the photons by the chromium and are raised to the n_2 level. As the energy cannot stay in this level, it drops back into the n_1 level which is metastable. The equivalent phonon energy of the n_2 - n_1 transitions are absorbed as heat. The transition from the metastable state, n_1, to the ground level commonly takes place spontaneously. The net result of our pumping the ruby populates the metastable state; this energy normally decays as it emits a fluorescent red light.

The principle of the LASER is embodied in the discovery that the energy transition from n_1 to the ground state can be stimulated by a powerful optical electromagnetic oscillation induced in the ruby. This wave motion increases the rate of the metastable

transition to the extent that the radiation, which normally lasts more than one millisecond, now is released in less than one microsecond. This corresponds to a power gain of some 1,000 times. In early LASERS the crystal was formed into a cavity structure by grinding the ends of the ruby cylinder flat and parallel. The ends were silvered so that very little of the fluorescent light could escape but, instead was reflected back and forth within the ruby in the manner of a Fabry and Perot interferometer. As the fluorescence increased and as the wave stimulated energy transitions built up, the optical standing wave within the ruby reached a power level of as much as 200 kilowatts. To get light out of the primative LASER, one end is given a partial silver coating which reflects 95% of the optical energy and transmits 5%. The release of the energy results in a series of pulses of light. This is shown to the right of Fig. 1, where the upper curve indicates the power of the pumping light and of the LASER as a function of time. You will notice that shortly after the pumping light has been on, that a series of pulses are emitted from the ruby. The initial LASER built at the Signal Corps Laboratory had a series of pulses with peak powers of only 600 watts. Finally, the lower right of the figure illustrates the output from a LASER which, by delayed stimulation, emits a single pulse of 3 megawatts power.

From the original Schawlow and Townes paper, the condition in which oscillations occur can be stated approximately as

$$N_1 - N_0 \geqslant K(1 - \alpha)\, AT^{\frac{1}{2}}$$

where α is the reflectivity of the ends, A is the area of the silvered ends and T is the temperature assuming a Boltzmann distribution of states. As we see from the equation the performance of the LASER is dependent upon the reflectivity of the ends, the area and the square root of the temperature. The pulse shaping illustrated in the figure has been achieved by changing the reflectivity during operation. To do this, one end of the ruby is cleaned of its silver coating and a rotating silver mirror placed at that end. When the mirror is not parallel to the end, α is small so that $1 - \alpha$ approaches 1 and the condition for oscillation approaches a limiting value. When the mirror rotates parallel to the end, oscillations can take place at a very small difference in states, for $1 - \alpha$ approaches zero. This then leads to the condition that the ruby is pumping during the time when a great number of n_1 energy states are needed for oscillation. As the pumping continues and the oscillation condition is approached the oscillation condition is drastically reduced by the rotating mirror as it swings parallel to the ruby face. The energy is "dumped" from the metastable state to give a peak pulse as shown in Fig. 2. The output energy as a function of time is also shown in Fig. 4. This discovery has made the construction of a range finder feasible.

MERRILL

LASER CHARACTERISTICS

The characteristics of the LASER which make it a unique source are: first, that the light is coherent and monochromatic; this is useful because all optical detectors are very wide band and detect many wavelengths at the same time. As you know, it is much easier to detect a light at night due to the dark background. Because the LASER is monochromatic, we can make the daytime background nearly black by placing a very narrow pass optical interference filter in the optical system. This almost eliminates the illuminated background and improves the detectivity. Second, the light emitted by the LASER is confined to a narrow beam, so narrow that at one mile, all the light emitted by the LASER falls within a circle less than 2 $\frac{1}{2}$ feet in diameter. Third, as mentioned before, the light is transmitted with very high peak power, which improves the signal to noise ratio.

A unique property associated with LASER output is connected with its coherence. The production of light, as we have known it, has always been associated with non-time coherent statistical processes, i.e. thermal excitation, fluorescence or phosphorescence and ionization. An electromagnetic standing wave in the LASER cavity enhances radiant emission which is induced in phase, and the radiation from the structure is coherent. Coherence may include two aspects: spatial coherence and temporal coherence. The degree of spatial coherence may be specified with respect to the degree of correlation of the LASER wave front. Spatial coherence exists if knowing the phase of one segment of a wave, the phase of all other segments can be described. Space coherence affects the optical properties of the wave and, when a wave is spatially coherent, it can be described completely by geometric and physical optics. It is achieved by a point source of light which can be made to exhibit the well known diffraction and interference phenomenon of optics. The LASER exhibits many of the properties of a point source.

One part of the face of the LASER output interferes with another, and the beam acts as a near plane wave with a very narrow beam. We have achieved beams in the Laboratory from the face of a $\frac{1}{4}$" ruby crystal which have a beam width of about 0.9 milliradians. If we compute the beam from an ideal source by the equation of the diameter of the first disc produced in Fraunhofer diffraction, $\theta = 1.22\lambda/D$, we find the ideal beam spread would be 0.01 milliradians. From this we see that the ruby is not very coherent. This is because the ruby has many imperfections and oscillates at different times in different modes which tend to disperse the wave.

Time coherence exists if knowing the phase of a wave at one time one can predict the phase of a wave at another time. Again this coherence is affected by the temperature and by the instantaneous dimensions of the ruby. The coherence is a function of the spectral line width.

The relation of the rate of delivering energy equal to the power has special significance depending upon the use of the LASER output. For detection purposes, it is important that we have very high peak power, but for most purposes it is important that a high average power be obtained. These relations can be illustrated by referring to the 3 megawatt peak power pulse which has a duration of about 10^{-7} seconds. The energy of this pulse is only 0.3 joule. If this is pulsed at the rate of six times in a minute the average power is only 30 milliwatts. An illustration of the output of a particular LASER is shown in Fig. 4 where the upper curve shows the output in the stationary state and the lower with a rotating mirror.

The efficiency of conversion is very low in all LASERS and especially in the ruby LASER. The pumping light, in the first place, must be coupled optically to the ruby. A reflector mirror is placed around the system in order to contain the energy in the cavity. There are optical losses at the reflecting walls and in reflections from the ruby surface. It may be possible to restrict these losses to 25%. The next serious loss is a spectral absorption loss. High energy sources of pumping light approach the quality of blackbody radiation, which has a broad wavelength spectrum, but the only useful portion is that wavelength band corresponding to the $n_2 - n_0$ transition of the chromium absorption lines at 5600A and 4100A. This is only about 10 per cent of the Xenon pump light. In the third place the energy corresponding to the phonon transition from the n_2 state to the n_1 metastable level of the ruby is absorbed in the structure and is lost as heat. This loss corresponds to about 30% of the remainder of the light. Thus we may expect an ultimate efficiency of some 5% of conversion. A LASER pumped by a helical FT 524 lamp at 1800 joules electrical input yields 3 joules with a stationary mirror and .3 joule with a rotating mirror and operated in the single pulse mode. These then operate with efficiencies of 0.2% and 0.02% at room temperature. It is interesting to note that the best lightweight capacitor that we have found has a weight of .018 lb per joule. In order to get a LASER output of only 100 joules at .2% efficiency we would need 900 lbs of capacitor; to get 10^5 joules output would require 450 tons if we used the same type of pump light and capacitors.

We can also consider efficiency from a conversion viewpoint. That is, in an optical system, our goal is to convert a light source into useful beam energy densities. In our example, we can consider the ratio of the energy per unit solid angle emitted by the LASER divided by the energy per unit solid angle emitted by our pump light. This gives us a conversion efficiency of 3×10^4. If we make this same conversion by the use of the same lamp at the focus of a 60" searchlight mirror we get a conversion efficiency of 600.

From simple geometrical considerations, the diameter of an image, i, formed with a beam having an angular diameter α and

focused by a lens of focal length, f, is $\alpha f = i$. If E, the energy per cm^2, is introduced into the lens through an aperture A then the energy density at the image becomes $E_i = 4EA/\pi\alpha^2 f^2$. As the energy density of the sun is .14 joule/cm^2/sec, the energy density of the sun's image formed on the retina would be about 120 joules/cm^2/sec. Eccles and Flynn[7] have been reported to find that sunlight of the order of 50 joules/cm^2/sec can cause retinal lesions. From this, it would appear that the eye cannot stand any more than a third of a second blink at the sun. In the case of the LASER the energy density of the beam deposited at the condensing lens is $E = 4\ E_o/\pi\alpha^2 R^2$ where E_o is the total energy emitted from the LASER and R is the distance from the LASER to the condensing lens. Then the energy per square centimeter delivered at the image is

$$E_i = 16\ E_o\ A/\pi^2\alpha^4 f^2 R^2$$

E_i is the very maximum amount of energy the eye can stand without damage. Then it is computed that the example LASER beam (.3 joule, 1 milliradian beam) will cause damage to the eye to distances as great as 300 meters. At this distance a tiny lesion would be produced as indicated in experiments with rabbits.

MATERIALS

As indicated before, there has been a considerable number of different materials which could be used for LASERS and in addition there are a very great number which could potentially be used. Some of the outstanding materials are listed in Table 1. These can be roughly classed as crystals, glasses and liquid solutions and gaseous LASERS. The ruby has been described predominantly throughout this paper because it has been most commonly used and has certain advantages leading to this preference. The output is in the visible region and it is possible to obtain the highest optical output of any other material. The ruby is also rugged and available in satisfactory shapes. Other crystals such as calcium fluoride and calcium tungstate have a complex output in the infrared and may require cooling. There is a problem in detecting the infrared with an adequate signal to noise and time response; indium antimonide has been used for this purpose. The gas LASER represented by the helium-neon mixture has advantages from a laboratory viewpoint as the energy states are not associated with the crystal lattice. It is possible to obtain a very pure spectrum and beam limited by diffraction; it is also easy to cool the gas LASER by providing a circulating gas. The disadvantage is that the concentration of atoms is very low and it is not expected that it will be possible to get a high power output.

TABLE 1

SOME MATERIALS COMMONLY USED IN LASERS

Active Element	Dominant Wavelength Emitted	Reference	Energy Emitted
$Al_2O_3: Cr^{+3}$ (pink ruby)	0.69 μ		20+ joules uncooled, 10+ megawatts peak, uncooled
$Al_2O_3:Cr^{+3}$ (pink ruby)	0.69 μ	(2)	cw 4 milliwatts (potential : 1 watt)
$CaF_2: U^{+3}$	2.56 μ	(3)	
$CaWO_4: Nd^{+3}$	1.056 μ	(4)	cw 1 milliwatt ac 1 milliwatt cooled
Glass Nd^{+3}	.88, 1.06, 1.34 μ	(5)	
He : Ne	1.15 μ (many other lines)	(6)	gas maser 1 milliwatt

USE OF THE LASER

As indicated earlier, the development of high peak power output has made the LASER attractive as a ranging device. The Ground Forces have always had a requirement for an accurate ground to ground range finder as they have had no range finder which was satisfactory when used with modern weapons. The optical range finders have been unsatisfactory because they depend mostly on a very small angle. Errors in this angle affect the range in proportion to the distance so that a practical range finder becomes too inaccurate for modern weapons. Small radar range finders have wide beams so that the range determination against a target is obscured by ground clutter. The LASER incorporates the advantages of the radar pulse method of determining range combined with a very small beam which can be sighted optically to clear unwanted targets. The operation of a LASER radar is illustrated in the diagram in Fig. 3. In the upper section the LASER transmitter emits a signal pulse. The pulse is detected by a solar cell which starts a time counter. The LASER beam traveling with the velocity of light is reflected from the target. The reflected light is detected by the receiver. The signal passes through an interference filter which cuts out

substantially other light such as that reflected by the sun from the target and meteorological aerosols. The photo cell detects the signal and stops the counter. The transmitted time as shown by the counter is calibrated with the light velocity so that the reading is directly in meters.

Potentially, the biggest use for the LASER is in connection with communications. A coherent light beam has a frequency of four orders of magnitude greater than a centimeter radio wave. It is interesting to think of the possibilities of the band width of an electromagnetic wave having a frequency of 2×10^{14} cycles per second. So far, this is not usable because there has been no detection means developed for these frequencies. If a satisfactory continuous wave LASER were available, we would have to settle for band width frequencies much smaller produced by conventional modulation. It is possible to modulate at microwave frequencies by optical rotation. Satisfactory communications depend on obtaining high C.W. power. Another serious limiting factor in communication is the attenuation of the signals by atmospheric conditions. It is possible to overcome this by transmitting the energy in a "Light Pipe." This "Light Pipe" can be used to carry the LASER beam with very small attenuation over large distances without relay amplifiers.

The use of LASER for illumination is only practical under special conditions. As indicated earlier in the paper, one of the basic advantages of the LASER is in connection with its ability to form narrow beams. For uses requiring a wide beam, it may be well to use more conventional light sources. However, in cases where point by point illumination, or illumination over very great distances, is required, the LASER may be attractive.

As the beam can be made very narrow, it is possible to intercept most of the beam with a practical receiver and collect substantially all the transmitted energy. The difficulty of transmitting power by the LASER is that it is an inefficient converter and as the wattage or energy is increased, the LASER, as we know it today, becomes very large, requires considerable power and wastes a large amount of heat.

The LASER beam may be concentrated by a lens into a very small area. With our example LASER, it is possible to obtain 30,000 joules per square centimeter on an area 1/10 millimeter in diameter at the focus of a 5 cm focal length lens. The reaction of this energy with materials can be very drastic as shown in Fig. 5. However, by controlling the output, it is possible to form a small hole in metals or to melt a very small area. This can be used to form holes in metal smaller than can be drilled, to etch away unwanted metals in micro-forming and to weld thin pieces of metal together. When properly controlled the LASER does not destroy the characteristics of the metal and can weld dissimilar metals.

The LASER promises to be an important tool for scientific investigations. It can be used in the laboratory where coherent light is required for the experiments in the measurement of optical and atomic transitions.

The LASER has also been studied for use in surgery with the eye. Experiments show that the LASER beam can be used as a "LASER retina coagulator." Experiments have been performed at Presbyterian Hospital, New York,[7] to form scar tissue for the retachment of partially or fully detached retinas to the supporting tissue. It can also be for purposes of destroying veins which furnish the blood supply to an eye tumor or to damaged tissues.

This paper has outlined some of the very high spots in the development of the optical LASER. The LASER has unique characteristics which make it of prime interest to scientific investigators. The application is still in its infancy and the most valuable employment of the LASER still depends on intensive research in the Laboratory. Some of the most useful employment depends upon new discoveries and employment of energy sources which have not been developed. The LASER is a real phenomenon which has an interesting application at the present time and which I am sure will overcome many of the problems that exist today.

REFERENCES

1. A. L. Schawlow, C. H. Townes, Phys. Rev., 29, pp 1940-1949, Dec. 1958.

2. D. F. Nelson, Elec. News, Jan 29, 1962.

3. H. A. Bostiek, J. R. Connor, Proc. IRE 50, p. 219 Feb. 1962.

4. L. F. Johnson, G. D. Boyd, K. Nassen, R. R. Soden, Proc. IRE 50, p. 213, Feb. 1962.

5. E. Snitzer, Phys. Rev., 7, p. 444, 1961.

6. Javan Bennett and Herritt, Phys. Res. Letters, vol. 6, pp 106-110, Feb. 1961.

7. M. M. Zaret, et al. RADC Technical Note 61-64 on contract AF 30(602)2215.

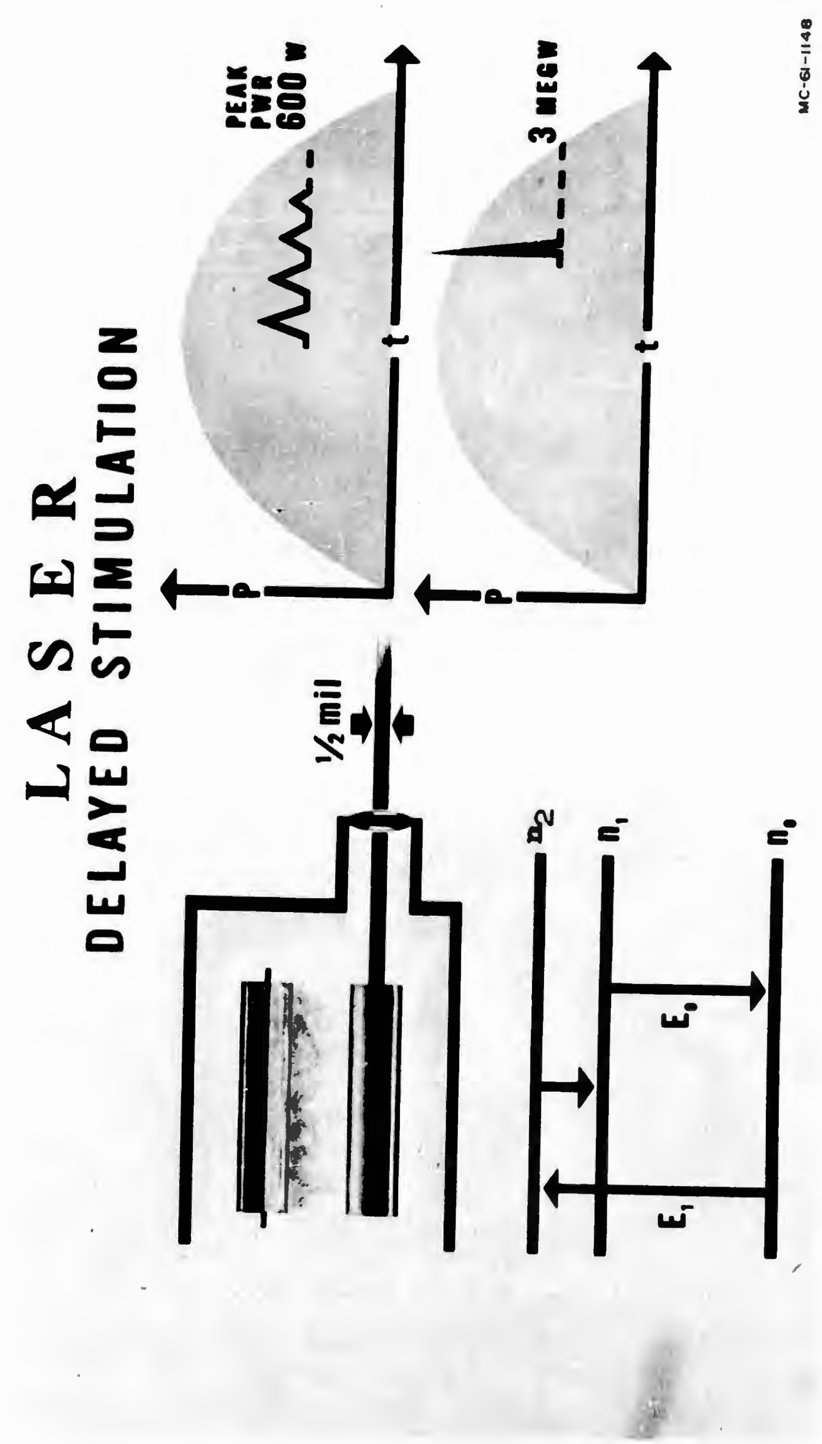
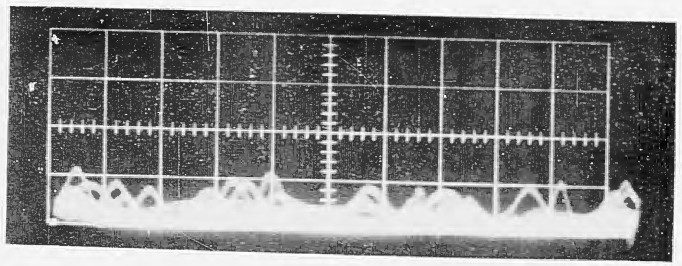

LASER
DELAYED STIMULATION
½ mil
P
t
PEAK
PWR
600 W
3 MEGW
n_2
n_1
n_0
E_0
E_1
MC-61-1148

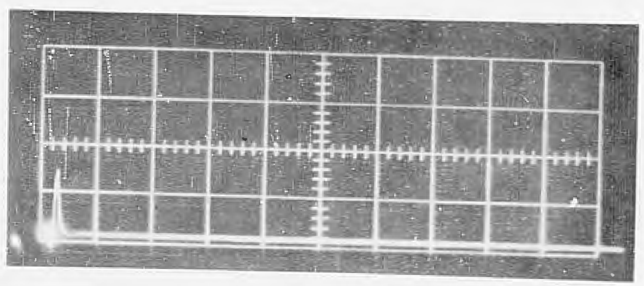
Fig. 1



VERTICAL SCALE: 500 WATTS PER DIV.

HORIZONTAL SCALE: 1 MICRO SECOND PER DIV.

NORMAL LASER OUTPUT, MAGNIFIED TIME SCALE



VERTICAL SCALE: 2 MEGAWATTS PER DIV.

HORIZONTAL SCALE: 1 MICRO SECOND PER DIV.

DELAYED STIMULATION LASER OUTPUT, MAGNIFIED TIME SCALE

Fig. 2

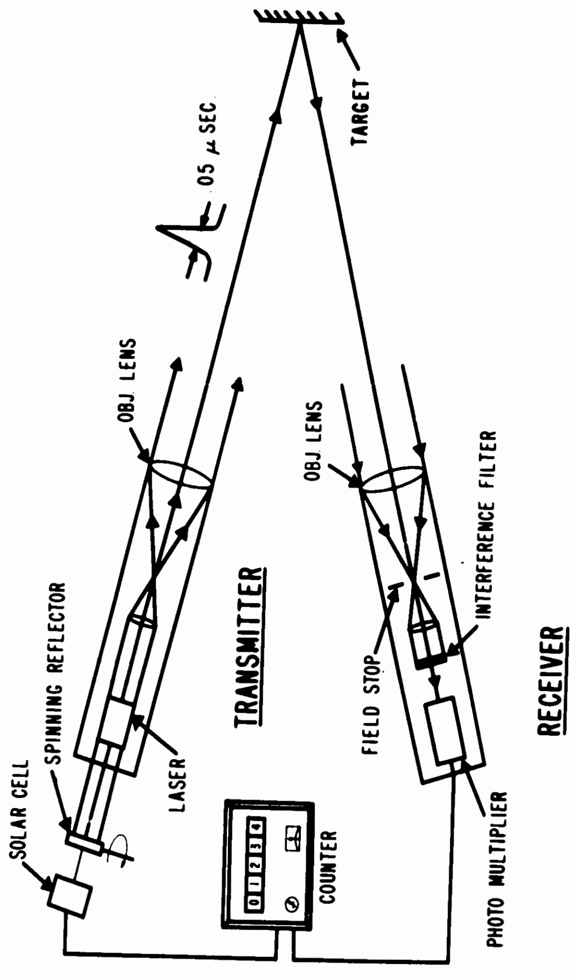


FIGURE 3 BLOCK DIAGRAM OF SINGLE - PULSE LASER RADAR

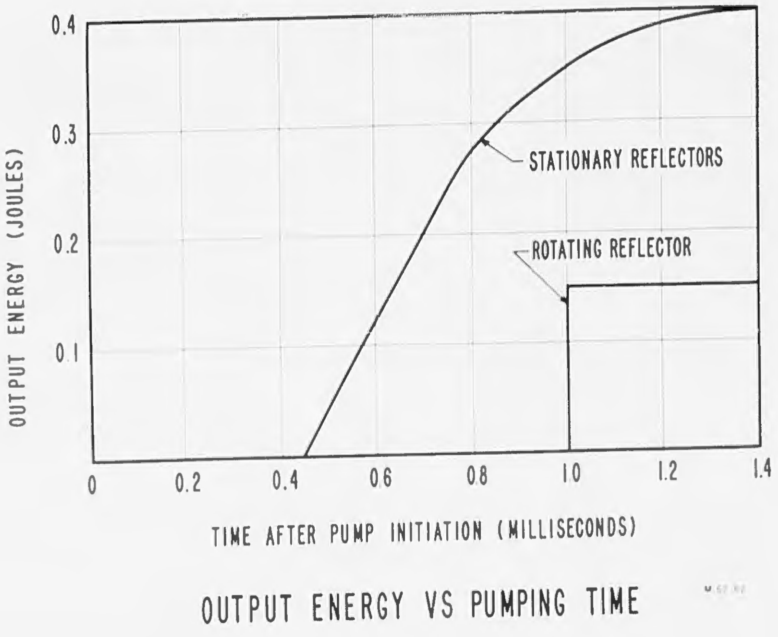


Fig. 4

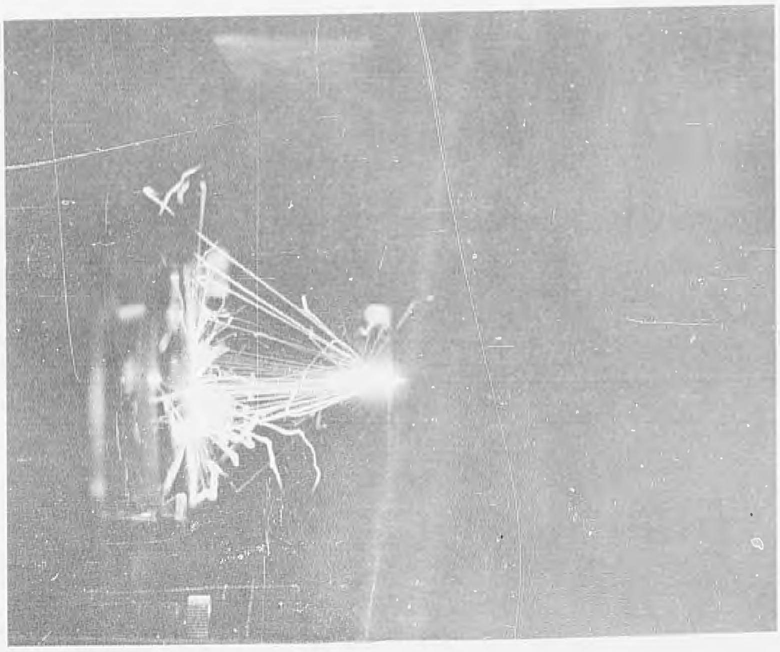


Fig. 5. LASER punching hole through 1/16" aluminum sheet.

REPLACING THPC IN THE APO-THPC FLAME-RESISTANT FINISH FOR TEXTILES

T. D. MILES, F. A. HOFFMAN, and A. MEROLA
QUARTERMASTER RESEARCH & ENGINEERING COMMAND
NATICK, MASS.

INTRODUCTION

Tris-1-aziridinyl phosphine oxide $(\overline{CH_2CH_2N})_3PO$ (1) in combination with tetrakis hydroxymethyl phosphonium chloride (7) is a flame retardant for cotton textiles. It appears that phosphorus compounds other than THPC can be used in combination with APO to impart flame resistance.

In studying the curing conditions for APO, a technique has been developed for the rapid screening of catalysts. A comparison of the infrared spectrum of APO before and after curing (Figure 1) shows that the absorption at 12.3 microns disappears upon curing.

To illustrate this, Figure 2 lists the absorption spectrum of a sample of APO which has been partially polymerized by air aging for three weeks as well as the spectra of the monomer, a commercial product (3), and that of a sample that has been completely polymerized at reduced pressure. Also shown for each sample is an estimation of the imine ring content by reaction with $Na_2S_2O_3$, which has been reported (8) as a semiquantitative method for the estimation of imine ring content according to the following reaction:

$$\overline{RNCH_2CH_2} + Na_2S_2O_3 + H_2O \xrightarrow{pH\ 5.2}$$

$$RNHCH_2CH_2S_2O_3Na + NaOH$$

The results listed in Figure 2 suggest that the change in absorption at 12.3 microns is due to the opening of the imine ring.

The polymerization of a ring compound usually proceeds by an interchange reaction, induced either catalytically or by the presence of small amounts of end-group-producing substances. For example,

water accelerates the polymerization of ε-caprolactam by hydrolysis of some of the compound followed by an interchange reaction.

$$\overline{CO(CH_2)_5NH} + H_2O \longrightarrow$$

$$HOCO(CH_2)_5NH_2 + \text{lactam} \longrightarrow$$

$$HOCO(CH_2)_5NHCO(CH_2)_5NH_2$$

Ethylene oxide polymerization may be initiated by alcohols, amines, and mercaptans capable of generating a hydroxyl group through reaction with the monomer (4). In the presence of strongly acidic or basic catalysts, successive addition of ethylene oxide molecules proceeds rapidly in the following manner:

$$\overline{CH_2CH_2O} + H + \longrightarrow HOCH_2CH_2 + \quad +$$

$$\overline{CH_2CH_2O} \longrightarrow HOCH_2CH_2OCH_2CH_2 +$$

Polymerization of cyclic compounds may also occur by ionic mechanisms under the influence of strong acids and in the absence of water and alcohol. Thus, in the presence of a strong acid or electron acceptor, ethylene oxide may polymerize violently.

From this it would be expected that compounds capable of attacking the nitrogen of APO would initiate polymerization such as a) alkylating agents capable of forming quaternary nitrogen compounds, b) acids, c) oxidizing agents, d) amines, e) inorganic salts, and f) Lewis acids. Compounds representing these various types were included in the screening.

EXPERIMENTAL

The compounds being screened as catalysts were applied in combination with APO to a viscose film which was dried at 212°F for five minutes and then cured at 285°F for five minutes. Infrared spectra of the treated films were obtained. (Disappearance of the absorption at 12.3 microns was used as the criterion for curing.) The following compounds were not catalysts for the polymerization of APO under the above conditions: boron trifluoride, diethylene triamine, ethylene diamine, hydrogen peroxide, methyl iodide, Monsanto catalyst AC, 2-amino-5-nitrobenzene sulfonic acid, phenol, succinic acid, boron trifluoride-triethanolamine complex. Table I lists the compounds which were found to be effective catalysts for APO under these conditions.

TRIS-1-AZIRIDINYL PHOSPHINE OXIDE -----A commercial product (3), an ethanol solution of APO, was used. Results of elemental

analysis after removal of ethanol are as follows: Calculated % C 41.6, Found 40.4; % H 6.99, Found 6.97; % N (Dumas) 24.36, Found 21.9; % P 17.9, Found 16.1.

Infrared spectra measurements were made with a Baird spectrophotometer. The sample was either ground into a K Br pellet (5,6) or run between two K Br pellets, 0.2 mm thick and 1.31 cm in diameter. A pellet holder (2) was constructed to hold the two K Br pellets together.

PREPARATION OF FILMS--------Unplasticized cellophane film was immersed in a 20% aqueous solution of tris-1-aziridinyl phosphine oxide containing from one to 15% of the catalyst being studied. Catalyst concentration was based on the resin solids content of the solution. Small samples of the cellulose film were allowed to stand in the solution for 30 minutes at room temperature to insure complete saturation. At the end of this period they were removed and blotted dry to remove the excess solution. Drying was accomplished by hanging the treated films in a circulating air oven for five minutes at 212°F. Curing was done in the same manner at 285°F for five minutes. Immediately after the samples were cured, the spectra of the films were obtained on a Baird spectrophotometer using an untreated film in the reference beam. (See Figure 1).

LABORATORY-SCALE APPLICATION-------The screening technique was followed first by laboratory and then by plant applications to a cotton fabric. Of the catalysts listed in Table I, diammonium phosphate has several advantages. It is among those that are effective in small amounts, it also has advantages of low cost and ease of handling. In addition, it is a flame retardant itself. For these reasons, diammonium phosphate was chosen for these trials.

The first trials, made on a laboratory padder (Butterworth three-roll pad), consisted of varying the bath concentration of APO and the diammonium phosphate. Pad liquor was made up to contain 5%, 10% and 15% APO. The catalyst concentrations for each of these baths ranged from 2½ to 15%; 0.3% Triton X-100 was used as a wetting agent. An 8.5-oz cotton sateen, nine inches wide, was given two dips, two nips, dried at 212°F for five minutes and cured at 300°F for five minutes.

Included in the trials was a bath containing 50% APO without any catalyst. The sateen fabric padded through this bath was dried for 15 minutes at 212°F and cured for five minutes at 300°F. This fabric after curing was not flame resistant.

Flame-resistance tests for the fabrics padded through the various concentrations of APC and DAP are shown in Table II. These results show that the diammonium phosphate is a catalyst for the curing of APC on the fabric as well as in the screening technique. Baths containing less than 15% APO did not fix sufficient resin on the fabric for flame resistance durable to laundering. With a bath containing 15% APO, the catalyst was effective from the 2½% level to the 15% DAP as shown by the durability of the flame resistance to laundering.

Of interest is whether the diammonium phosphate, in addition to being a catalyst for APO, contributes to the flame resistance of the treated fabric, i.e., whether any of the phosphorus from the diammonium phosphate is retained in the treated fabric. Since both APO and diammonium phosphate contain phosphorus, a sample of diammonium phosphate containing radioactive phosphorus (p^{32}) was prepared.

An 8.5 oz. cotton sateen fabric was immersed in a bath containing 5% diammonium phosphate, the phosphorus component of which was radioactive (p^{32}). The fabric was squeezed, oven dried at 100°C and cured for five minutes at 140°C. Using a beta-particle counter, it was found that only 6.9% of the radioactive diammonium phosphate was retained after four cold water rinses.

The same type of cotton fabric was then immersed in a bath containing 5% radioactive diammonium phosphate and 18% APO. The fabric was oven dried at 100°C and cured for five minutes at 140°C. After four cold water rinses the fabric was found by the beta-particle counter to contain 43.6% of the original radioactive diammonium phosphate. After boiling for one-half hour in soap and soda ash (0.2% soap and 0.1% soda ash), the fabric was found by the same technique to have retained 36.6% of the original radioactive diammonium phosphate. These results are summarized in Figure 3.

PILOT-PLANT APPLICATION--------Pilot-plant applications were made on QM equipment which consisted of a Gessner two-bowl padder, a pin tenter inclosed with a flue-type drying oven and a curing oven. Finishing procedures were not continuous and the fabric was batched up after each operation. Since the equipment had only one squeeze pressure, the wet pickup could not be varied. Higher wet pickups in the range of 70% were obtained on coating-type equipment having adjustable squeeze rolls. Fifty yards of Cloth, Sateen, 8.5 oz., (MIL-C-12095) was padded through a bath containing 5% diammonium phosphate and 18.2% APO in the following bath formulation:

	4.25 lbs	diammonium phosphate
	21.25	Imine IP (based on 73% solids)
	0.4	Triton X-100
	59.1	water
Total	85 lbs	(10 gal)

The wet pickup was 75%. The fabric was dried at 250°F for two minutes and cured at 300°F for five minutes. Table III lists the performance test results for this fabric.

With the equipment that gave a lower wet pickup, i.e., 55%, the finish was applied in a two-pass process. The first pass was through a bath containing 15% APO and 10% DAP. After drying and curing, the fabric was passed through a bath containing 8.75% APO and 7% DAP. The fabric was again dried at 250°F for two minutes and cured at 300°F for five minutes. Table IV lists the performance test results for this fabric. There is also evidence that the bath stability of DAPO is greater than APO-THPC. Samples taken from an APO-THPC bath at room temperature polymerize into a solid on standing overnight. Samples of DAPO bath have stood for at least two weeks without visual evidence of polymerization.

CONCLUSIONS

The general problem of what types of chemical compounds can be combined with tris-1-aziridinyl phosphine oxide (APO) to produce durable flame retardants for cellulosic fabrics has been explored. APO by itself is not polymerized into a durable flame retardant within the time and temperature limits set by conventional textile processing equipment. Based upon the changes in the infrared spectrum of APO during polymerization a method for the rapid screening of candidate catalysts was developed. Of the wide variety of chemicals screened, compounds that are acidic under curing conditions were found to be the most effective catalysts.

One of these catalysts, diammonium phosphate, combined with APO was examined in laboratory and pilot-plant applications. A radioactive tracer study showed that approximately one-half of the diammonium phosphate is retained in the fabric after washing and thus contributes to the flame resistance as well as serving as a catalyst. The combination has several advantages over APO-THPC. One is a reduction in chemical cost. Another is that there is no release of formaldehyde fumes in the pad bath, drying oven or curing oven. This makes DAPO a more acceptable formulation from a plant application standpoint, particularly where ventilation is limited. Another advantage is greater bath stability. Thus the combination of diammonium phosphate and APO provides a new durable flame retardant for cotton.

ACKNOWLEDGMENT

The authors acknowledge the assistance of Raymond D. Cooper and Robert C. Grubinskas of the Pioneering Research Division, QM R&E Command in the work involving the use of radioactive materials.

REFERENCES

(1) Bestian, H, Ann 566, 210 (1950).
(2) Feazel, C. E, and Vershot, E. A, J Polymer Sci 25, 351-3 (1957).
(3) Farbwerke Hoechst, Frankfurt AM, Germany.
(4) Flory, P. J, "Principles of Polymer Chemistry" 57-61, Cornell University Press, Ithaca, N.Y. (1953)
(5) Harp, W. R, Stone, H, and Otvos, J. W, "Quantitative Absorption Spectrometry of Solids", Pittsburgh Conference on Anal Chem and Applied Spectroscopy, Mar. 1954.
(6) Perkin-Elmer Instrument News 4, 1 (1953).
(7) Reeves, W. A, Drake, G. L, Chance, L. H, and Guthrie, J. D, Textile Research J 27, No. 3 260-6 (1957).
(8) Schafer, F. C, Geoghigan, J. T, and Kaiser, D. W, J Am Chem Soc 77, 5918 (1955).
(9) Drake, G. L, and Guthrie, J. D, Textile Research J 29, 155 (1959).

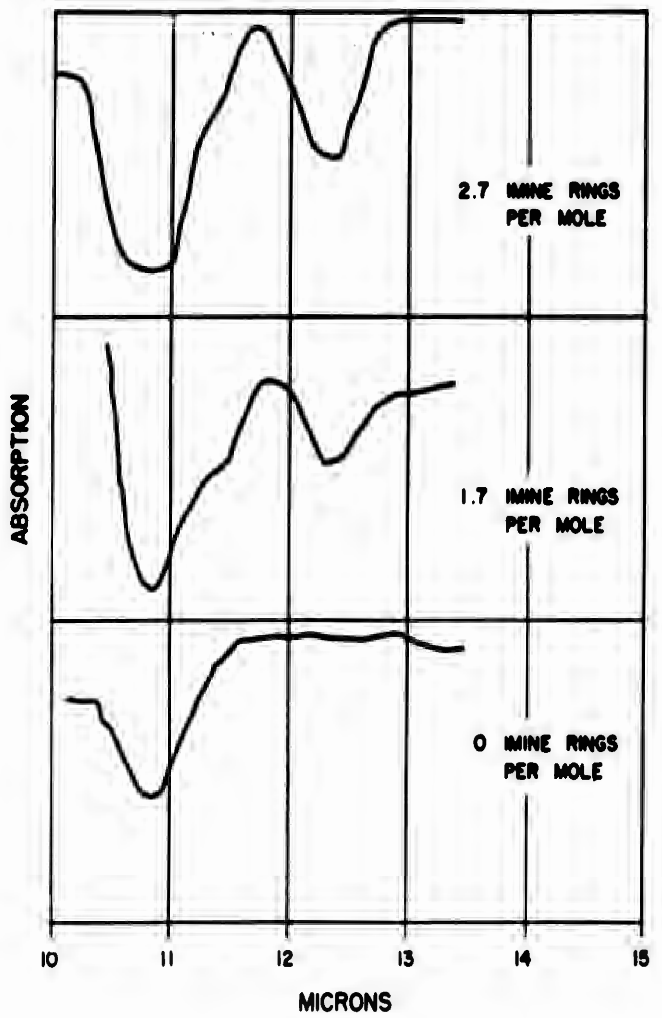
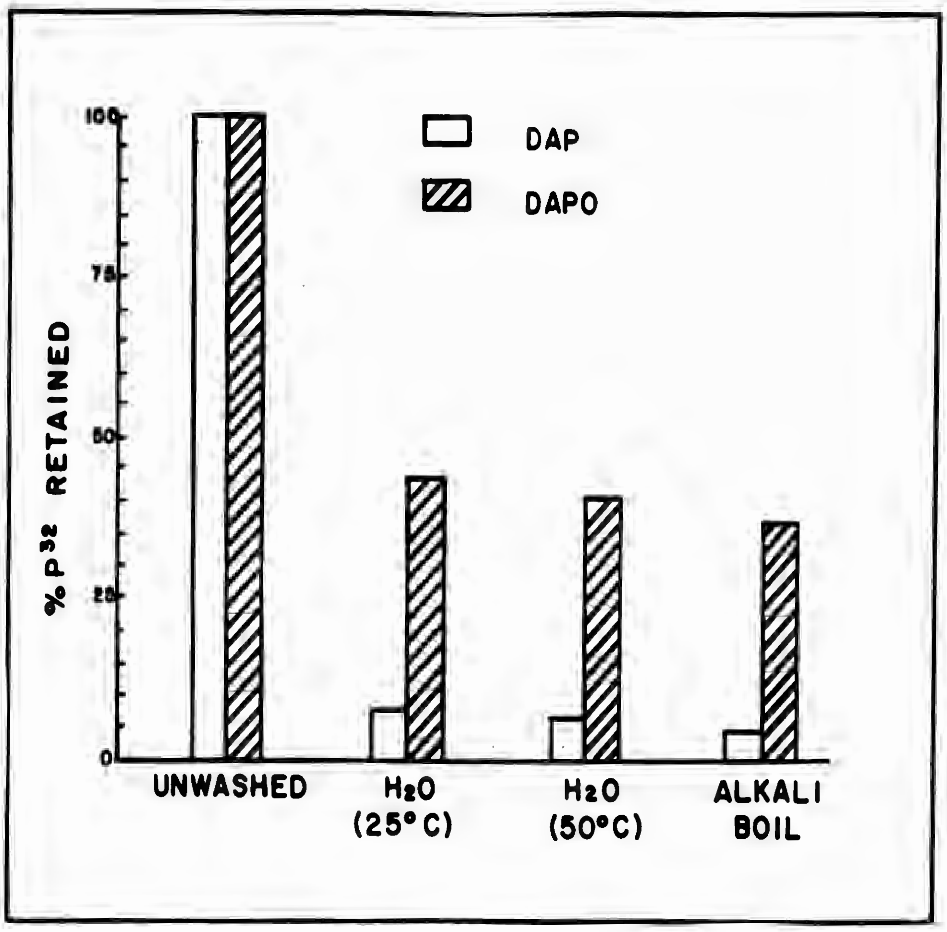

2.7 IMINE RINGS PER MOLE
1.7 IMINE RINGS PER MOLE
0 IMINE RINGS PER MOLE
ABSORPTION
10
11
12
13
14
15
MICRONS

Figure 2



100
75
50
25
0
% P32 RETAINED
DAP
DAPO
UNWASHED
H2O (25° C)
H2O (50°C)
ALKALI BOIL

Fig. III

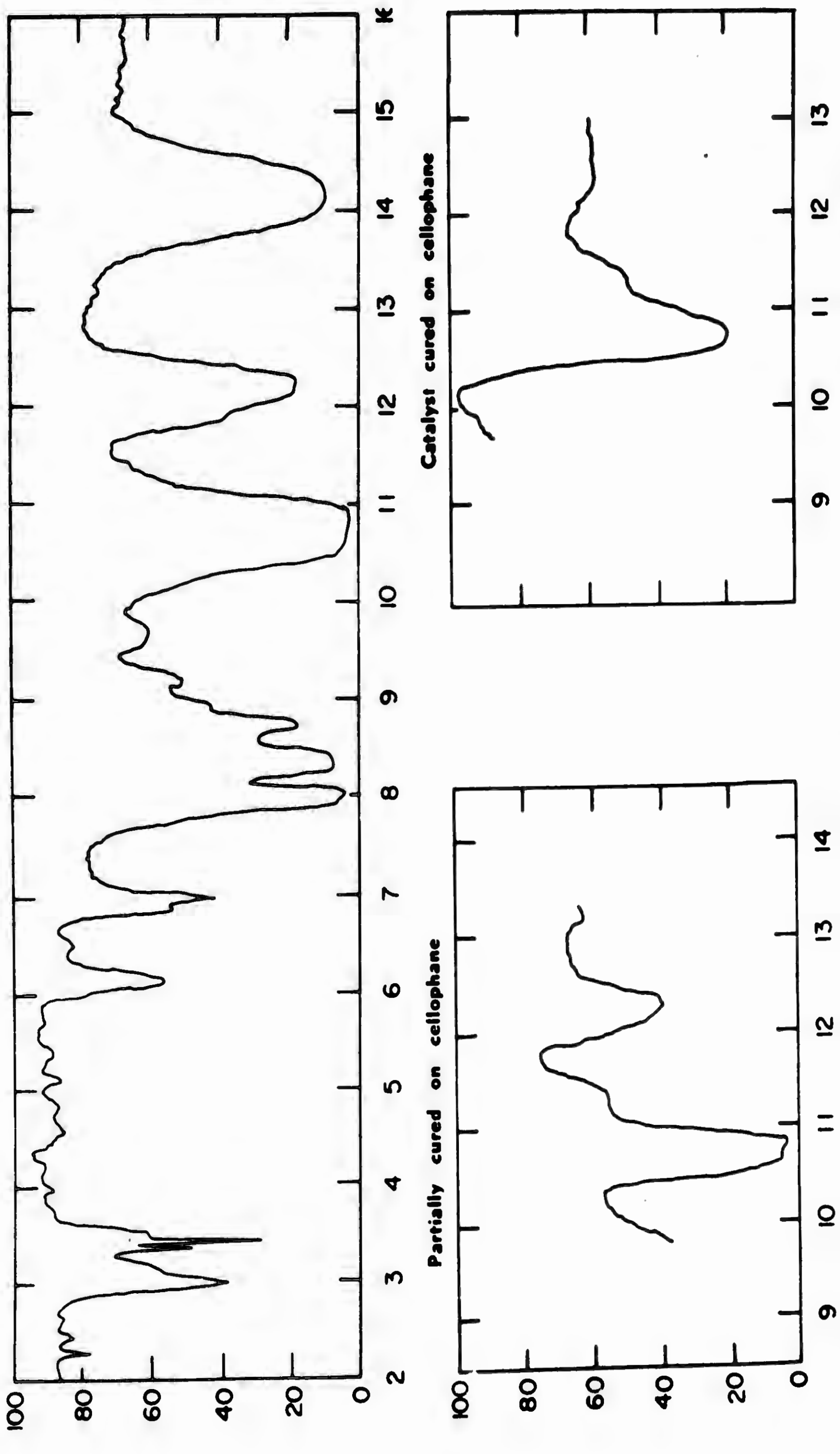

Catalyst cured on cellophane
Partially cured on cellophane
100
80
60
40
20
0
2
3
4
5
6
7
8
9
10
11
12
13
14
15
μ

Figure 1

TABLE II

Effect of varying bath concentrations of APO and DAP on flame resistance of 8.5-oz sateen

	5% APO 5% DAP	5% APO 10% DAP	10% APO 5% DAP	10% APO 10% DAP	15% APO 2½% DAP	15% APO 5% DAP	15% APO 10% DAP	15% APO 15% DAP
Wet pickup (%)	63.0	64.6	66.0	65.5	63.0	63.5	68.3	66.0
Resin add-on (dry weight)	4.4	6.7	7.0	10.5	7.6	10.5	14.8	13.7
Initial	0-1-5.9*	0-1-4.9	0-1-4.9	0-1-3.9	0-1-4.2	0-1-3.8	0-1-3.9	0-1-2.7
Laundered:								
one cycle	BE**	BE	0-1-4.4	0-1-5.0	0-2-4.7	0-1-3.9	0-1-4.0	0-1.7-3.1
three cycles	BE	BE	0-1-4.3	2-1-5.3	0-1.3-4.5	0-1-3.8	0-1-4.4	0-1.1-4.0
six cycles	—	—	4-0-7.8	BE	0-1.2-5.3	0-1-4.5	0-1-4.7	0-1.1-4.3
nine cycles	—	—	BE	0-1-4.7	0-1.2-5.0	0-1-4.5	0-1-4.7	0-1.4-4.1
12 cycles	—	—	BE	4-1-5.8	0-2.1-6.0	0-1-4.4	0-1-4.4	0-1.0-4.0
15 cycles	—	—	—	BE	0-1.2-4.3	0-1-4.4	0-1.7-4.4	0-1.2-3.7

*Vertical Bunsen—Method 5902 (*10*) (afterflame (sec), char length (in))

**BE—Burned to end

TABLE I

Catalysts for the polymerization of tris-1-aziridinyl phosphine oxide showing the lowest effective concentration (%)

	2	4	5	6	8	10	15
Oxalic acid					x		
Sulphonic acid			x				
Formic acid			x				
Sulfuric acid			x				
Phosphoric acid					x		
Hypophosphorus acid				x			
Phenol disulfonic acid	x						
Zinc perfluoborate (9)	x						
Ammonium chloride							x
Diammonium phosphate	x						
Urea phosphate						x	

TABLE III

Strength and flame resistance of Cloth, Cotton, Sateen, 8.5-oz, DAPO treated (18.2% APO—5% DAP bath)

		Tear strength (wxf)	Flame tests AF	AG	CL	% P
Untreated		10.4 × 8.2				
Treated	Initial	4.9 × 3.6	0	1.2	3.8	
	Laundered—one cycle	4.7 × 3.3	0	1.4	4.0	2.24
	three cycles	4.2 × 3.1	0	1.4	4.5	
	six cycles	4.6 × 3.8	0	1.4	4.7	
	nine cycles	4.2 × 3.5	0	1.3	4.7	
	12 cycles	4.4 × 3.3	0	1.3	4.7	
	15 cycles	4.1 × 3.8	0	1.2	5.0	

EVALUATION OF HYPERSONIC PROBE MATERIALS IN A WATER-STABILIZED ELECTRIC ARC AT TEMPERATURES TO 3200°C (5795°F) FOR 60 SECONDS

D. J. MOLELLA
PICATINNY ARSENAL
DOVER, NEW JERSEY

INTRODUCTION:

The probe configuration evaluated was that which was proposed for possible use on a theoretical Mach 8.5 missile. The pressure sensing probe system must have the capability of measuring total and static pressure for velocity determination. From the known velocity, the maximum allowable movement of the aerodynamic control surfaces are determined so that the maneuvering "G" limits of the missile are not exceeded. The probe also performs certain warhead functions.

Numerous experimental probes were subjected to wind tunnel and simulated flight tests in order to determine the most suitable aerodynamic design for this probe. In this design aerodynamic heating must also be considered since the maximum Mach number will be 8.5. When such Mach numbers are involved, it is customary to ascertain the trajectory that yields the greatest aerodynamic missile heat input and then design the system to meet the heat requirements of this trajectory. Such a trajectory is called the "Thermodynamic Design Trajectory".

From this design trajectory, which results in maximum heating, heat fluxes expected at the forward portion of the probes to be tested were determined (1). These flux calculations were based on the assumption of a cold wall. Since it was known that the flow field of the electric arc was subsonic, the flow away from the stagnation point would not be similar to that which exists when supersonic flow is present on the probe during flight. However, the flow at the stagnation point in the arc subsonic flow field was similar to the flow field present on the probe during flight. From these assumptions and calculations a heat flux vs time testing program on the probe models was proposed, as follows:

initial input - 65 BTU/ft^2-sec
maximum input - 450 BTU/ft^2-sec to be reached in 25 seconds
final input - 225 BTU/ft^2-sec at the end of 60 seconds

If a probe material or a material with a protective coating could withstand the maximum service heat inputs without ablating, oxidizing, burning or softening, the probe geometric configuration would be maintained throughout the trajectory and the probe would then perform its functions satisfactorily. In order to assure maintenance of the geometric configuration, only materials having high thermal conductivities, heat capacities and mechanical strengths at low and elevated temperatures should be considered. In the final selection of the best material, availability, fabricability and cost of the material should also be considered. This investigation was conducted to evaluate various materials and coatings to find a suitable combination to fulfill the above requirements.

Since a water-stabilized electric arc can simulate the heat flux vs time that are obtained from the maximum trajectory, this apparatus was used for determining the most suitable probe material.

SYMBOLS:

Mo	molybdenum
$90W+10ZrO_2$	90% tungsten + 10% zirconia
Ti	titanium
Al-Cr-Si	aluminum-chromium-silicon
$MoSi_2$	molybdenum disilicide
ZrB_2	zirconium diboride
Si C	silicon carbide
W C	tungsten carbide
Co	cobalt
Cb-Ta-Zr	columbium-tantalum-zirconium

APPARATUS AND EXPERIMENTAL PROCEDURE:

HEAT TESTS:

A water-stabilized electric arc was used in the evaluation of the materials for resistance to high temperature gas flow. The arc operates in the vortex formed by a rapidly swirling body of water. Extremely high temperatures up to 25,000°F and high heat inputs (2500 BTU/ft^2-sec) are developed due to the increased current density arising from the restricted cross sectional area of the vortex and the cooling of the outer arc regions, which forces most of the current to flow in the center (2). The arc is struck in a vortex of water between a ½" dia graphite disk with a 3/16" dia hole in the center as the cathode. The cathode is mounted in a water-cooled horizontal metal plate at ground potential, while the anode is vertically held by a collet directly below the cathode cavity. Figure 1 shows a schematic diagram of this apparatus. A lucite "swirl chamber" surrounds the anode. During operation water is introduced tangentially into the chamber at 36 psi and forms the vortex in which the arc burns.

The power supply for the arc consists of four welding generators connected in series. For stable operation about one half of

the power is dissipated by a ballast resistor connected in series with the arc. Constant arc voltage is maintained by controlling a hydraulic mechanism which feeds the lower electrode at the proper rate by means of a voltage sensing device connected across the arc. Under these conditions, a power dissipation of 33KW results at the arc. A water-cooled parabolic cathode nozzle was developed for testing the probes in order to provide optimum reproducibility in the arc environment.

The models were mounted on an adjustable support and positioned so that when swung into the arc, the heat flux on the model simulated approximately the heat flux expected within a second or so of zero time (missile at rest). Figure 2 shows the water-stabilized electric arc apparatus before operation or prior to swinging the model into the jet. The power was then turned on and the arc flow field was set to the same voltage and amperage that was used to determine the steady temperature that would occur on the probe tip vs distance from the nozzle exit. These calibrations were previously made by using a black body (graphite) probe. When stable arc operation was attained, the probe model was positioned in the flow field. It was then moved axially towards the arc nozzle so that the heat flux vs time simulated the thermodynamic design trajectory and so that the maximum heat flux input (450 BTU/ft^2-sec) on the model occurred at 25 seconds. Figure 3 shows the electric arc in operation with a model under test.

High speed color motion pictures were made of the entire test run to show the heat input effects on the models.

PROBE MODELS:

The materials selected for evaluation on the models were based on current availability and fabricability and were also based on the results of a previous investigation of available high temperature materials (3). The materials and coatings selected for evaluation are shown in Table 1. All models were fabricated to the probe configurations (Figures 2, 4 and 5) at Picatinny Arsenal. The models had a concave nose-cylinder geometric shape with a pressure cavity at the nose. Those to be coated were sent to private concerns for application of the various oxidation-resistant coatings. Identical coatings were also applied to flat specimens so that the characteristics of the coatings on the models prior to the heat tests could be determined. All of the coated and uncoated models and the coated flat specimens were ground to a RMS-32 surface finish.

Some of the coatings, listed below, were proprietary; therefore, very little was known about their compositions.

W-2 - Chromalloy Corp vapor-deposited coating on molybdenum believed to consist mostly of chromium along with some aluminum and silicon.

DURAK-MG - Chromizing Corp coating believed to be similar to the W-2 coating.

LM-5 - Linde Co flame-sprayed, modified molybdenum disilicide coating on molybdenum with additions of chromium carbide and aluminum.

Gradated Zirconia - Linde Co flame-sprayed, multi-layer coating on molybdenum consisting of a metal-rich+zirconia undercoat and zirconia-rich+metal overcoat.

Am. Lava - American Lava Corp coating on graphite believed to consist of tungsten and silicon carbide.

SURFACE TEMPERATURE DETERMINATIONS:

A calibrated optical pyrometer was used to determine the highest surface temperature that was attained by each model during the electric arc heat test. The time (seconds) when this maximum temperature was reached was recorded.

Caution must be exercised in interpreting the optical pyrometer observations (4). The pyrometer measures a luminescence temperature, T_S, which is related to the true temperature, T, by the equation, as follows:

$$\frac{1}{T} - \frac{1}{T_S} = \frac{\lambda \log_e \epsilon_\lambda}{C_2}$$

where ϵ_λ is the spectral emissivity at wave length, λ (microns) and C_2 is the constant in the Wien equation for distribution of energy in the spectrum and has a value of 1.438 cm degrees. It was decided that correcting the optical pyrometer readings for emissivity differences was unnecessary since the same heat input flow field was repeated for each heat test and therefore, subjected the probes to the maximum heating trajectory. Also, little or no data were available on the emissivities of the various coatings and base materials used.

METALLURGICAL EXAMINATIONS:

Visual examinations were made on the models before and after subjecting them to the arc heat tests in order to determine the extent of damage, if any, to the surfaces. The naked eye and standard low power (below 10X) optical magnifiers were used for these inspections. Still photographs were taken before and after heat testing the models.

Microscopic examinations were conducted on the flat specimens to determine the metallurgical characteristics of the coating before the arc tests. Microscopic examinations were also conducted on all the models after the heat tests to determine the effects of the heat inputs. These examinations were made at magnifications up to 1000 times on polished specimens cut from the flat sheets and at the nose portions of the probes. Standard metallographic procedures and microscopes were used. Photomicrographs showing the heat effects on the coatings and base materials were taken.

RESULTS AND DISCUSSION:

WATER-STABILIZED ELECTRIC ARC TESTS:

Descriptions of the various materials and coatings used on the probe models and the heat effects on them are shown in Table 1. These effects were determined by viewing color motion pictures of the models taken during the electric arc tests. Also included in the table are the maximum surface temperatures attained by the models, as determined with a calibrated optical pyrometer, and the time (seconds) to reach these temperatures.

TABLE 1. Optical Pyrometer and Motion Picture Observation of Probes During Electric Arc Heat Tests

Probe Material	Max Observed Temp (°C)	Time to Reach Max Temp (sec)	Color Motion Picture Observations
Mo with $90W+10ZrO_2$ flame-sprayed coating, .015" thick	2030	35	Excellent preservation of probe tip contour throughout test
Mo with Al-Cr-Si flame-sprayed coating, .010" thick	1790	35	Excellent preservation of probe tip contour throughout test
Mo with gradated Zirconia flame-sprayed coating, .015" thick	2180	40	Flaking started immediately but protection was very good
Mo with flame-sprayed ZrB_2 coating, .010" thick	2325	40	Coating flaked at 15 sec. Slight erosion and oxidation apparent at 35 sec.
Mo + .5 Ti with vapor-deposited coating, .0014" thick	1970	35	Excellent preservation of probe tip contour throughout test
Mo with $MoSi_2$ vapor-deposited coating, .0014" thick	2155	35	Coating failed at 20 sec. Erosion and oxidation apparent

TABLE 1. (Continued)

Probe Material	Max Observed Temp (^{o}C)	Time to Reach Max Temp (sec)	Color Motion Picture Observations
Mo with DURAK-MG vapor-deposited coating, .0014" thick	2120	55	Coating failure at 34 sec. Leading edge exposed at 40 sec.
Mo with DURAK-MG vapor-deposited coating, .0014" thick	2160	35	Coating failure at 20 sec. some Mo oxidation apparent
Mo with Al-Si flame-sprayed, oxidized coating, .010" thick	2160	35	Coating failure at 28 sec. with oxidation of Mo apparent
Mo with W-2 vapor-deposited coating, .0014" thick	1880	37	Full size probe. Excellent preservation of probe tip contour throughout test
Mo with W-2 vapor-deposited coating, .0014" thick	2180	40	Coating failure at 22 sec. Catastrophic oxidation of Mo apparent at 27 sec.
Mo with LM-5 flame-sprayed coating, .015" thick	2110	32	Rapid flaking of coating at 22 sec. Catastrophic oxidation of Mo evident
Graphite with SiC vapor-deposited coating, .010" thick	1840	38	Excellent preservation of probe tip contour throughout test
Graphite with Am. Lava coating, .005" thick	2075	35	Slight oxidation and erosion apparent at sec.
Graphite with 90W+10ZrO_2 flame-sprayed coating, .015" thick	2260	35	Coating failure on rim at 25 sec. Slight erosion apparent

TABLE 1. (Continued)

Probe Material	Max Observed Temp (°C)	Time to Reach Max Temp (sec)	Color Motion Picture Observations
Uncoated tungsten	1945	25	Excellent preservation of contour throughout test. Slight oxidation vapors apparent
Uncoated, self-bonded silicon carbide	2180	35	Very slight erosion and oxidation at 22 sec.
Uncoated graphite	2115	30	Moderate erosions apparent throughout test. Model fractured at threads when swung out of arc
Uncoated ZrB_2	3200+	20	Moderate erosion throughout test. Observed temp beyond range of pyrometer
Uncoated molybdenum	2135	30	Catastrophic oxidation apparent at 25 sec.
Uncoated WC with Co binder	2200	30	Melting and bubbling out of binder apparent at 24 sec.
Uncoated, porous MgO with resin binder	2160	26	Cracking and erosion at 21 sec. Ablation apparent throughout test
Uncoated Cb-Ta-Zr Alloy	2305	30	Melting evident at 14 sec. Violent spattering of molten metal at 23 sec, when test terminated

The above results show that one model with a W-2 coating and another with a $MoSi_2$ coating withstood the heat inputs well; whereas, other probes having identical coatings exhibited poor heat resistance. These discrepancies in performance were attributed to

differences in quality or workmanship of the coatings. Other coatings that failed, stripped or flaked-off within 35 seconds as a result of brittleness or poor erosion and thermal shock resistance. The main causes for poor heat resistances of the uncoated models were melting, erosion and oxidation.

Although subjected to the same heat inputs, the maximum surface temperatures at the tips of the models varied considerably. These variances were due to the different emissivities, decomposition rates, densities, thermal conductivities and heat capacities of the various materials and coatings. An uncoated zirconium diboride (ZrB_2) model attained the highest surface temperature, above 3200°C (5795°F), in the least amount of time (20 sec) due to lower thermal conductivity and less heat capacity compared with most of the other materials.

METALLURGICAL EXAMINATIONS:

Heat effects on the electric arc-tested probes, which were determined by visual or macroscopic examinations at magnifications up to 10 times, are shown in Figures 4 and 5. The results from these examinations and subsequent microscopic examinations at magnifications up to 1000 times corroborated the results previously disclosed by observations of the color motion pictures taken during the electric arc heat tests. These results are shown below.

COATED MOLYBDENUM PROBES:

Out of twelve oxidation-resistant coatings tested on molybdenum, the first probe model, Figure 4, which had a 90% tungsten + 10% zirconia coating exhibited the least damage. Subsequent microscopic examination disclosed that, except for some slight oxidation and erosion, this coating had virtually remained intact with no apparent damage to the base metal, as shown on the photomicrographs in Figure 6. The heat effects on the structure of this same coating on graphite is shown in Figure 7. The next four molybdenum probes (Fig 4) containing aluminum-chromium-silicon, gradated zirconia, zirconium and molybdenum disilicide coatings exhibited about the same degree of damage. These probes were slightly eroded and some oxidation products in powder form had redeposited at the rear. Even though the gradated zirconia coating on the molybdenum was cracked and had flaked-off somewhat at the leading edge radius, this model had still maintained its basic configuration. Subsequent microscopic examination revealed that the metal-rich undercoat remained intact and evidently was able to protect the base metal from catastrophic oxidation.

As previously mentioned, one molybdenum probe with a molybdenum disilicide ($MoSi_2$) coating and another full-size probe with a W-2 coating performed well. On the other hand, another probe model with an identical $MoSi_2$ coating exhibited a considerable amount of scaling at the nose tip and pressure cavity. The other model with an identical W-2 coating as the full-size probe showed visible evidences

of scaling and catastrophic oxidation. These discrepancies were evidently due to differences in the quality or workmanship of the coatings. According to the findings of others (5), the capability of a coating is dependent on its quality. The coatings which performed poorly were found to be blistered and contained numerous microcracks prior to being subjected to the heat tests. These defects were found by examining flat specimens having these identical coatings. No microscopic examination was made on the full-size probe with the W-2 coating which performed well. This probe was coated at a later date than the probe model which performed poorly and therefore, probably did not contain any microcracks or other detrimental defects. This improvement was most likely due to optimization in the quality or workmanship of the W-2 coating (5). $MoSi_2$ and W-2 coatings are vapor-deposited coatings that contain certain amounts of silicon. Other molybdenum probes with vapor-deposited and/or silicon-bearing coatings, such as DURAK-MG, aluminum-silicon (AL-Si) and LM-5 coatings also exhibited visible evidences of scaling at the nose tips and pressure cavities. No microcracks were visible upon microscopic examination of flat specimens having these identical coatings. Since it was impossible to examine all areas of the coatings microscopically, microcracks could still have been present in the coatings before the heat tests. Thermal cracks could develop during the heat tests which would cause subsequent scaling and oxidation of the exposed surface of the base metal. In addition to the scaling, the model with the LM-5 had oxidized catastrophically. This was attributed to the poor mechanical bond between the coating and the base metal, as shown in Figure 8.

COATED GRAPHITE PROBES:

Metallurgical examination of three graphite probes with oxidation-resistant coatings disclosed that the probe with the self-bonded, silicon carbide (SiC) coating exhibited the least damage (slight erosion), as shown in Figure 5. This model, however, fractured when accidently dropped during transportation. Microscopic examination revealed that this probe was virtually unaffected by the heat test. Due to the extreme differences in the hardnesses between the coating and the graphite, it was not possible to prepare suitable photomicrographs. The other two models containing the 90% tungsten + 10% zirconia (Fig 7) and the Am Lava coatings exhibited about the same degree of erosion and oxidation damage. As previously mentioned, the graphite probe with the $90W + 10ZrO_2$ coating broke at the threads when swung out of the arc. This susceptibility to fracturing at such low stress values indicates that graphite, by itself, is too brittle at low temperatures to be a suitable, basic probe material.

UNCOATED PROBES:

Out of eight uncoated materials arc-tested, tungsten exhibited the least damage (Fig 5). This probe oxidized; however, the oxidation was very uniform and the nose tip receded (ablated)

slightly. Self-bonded silicon carbide oxidized and eroded slightly. Uncoated graphite had moderately oxidized and eroded. Uncoated zirdonium diboride (ZrB_2) was moderately oxidized and eroded and contained thermal cracks. As expected, uncoated molybdenum had oxidized catastrophically and the nose had receded about 0.17 inch. Severe deformation and degradation of the nose tip and pressure cavity were evident upon examination of the tungsten carbide (WC) containing the cobalt binder as a result of melting of the binder. Numerous thermal cracks were located throughout the uncoated probe model made of porous magnesia (MgO) with a phenolic resin binder. Charring and deformation of the nose was also evident. About 0.2 inch of the nose of the uncoated columbium-tantalum-zirconium (Cb-Ta-Zr) probe had receded or melted away. The pressure cavity was almost completely filled with re-melted and solidified metal.

CONCLUSIONS:

1. Molybdenum when adequately protected against oxidation was found to be superior to the other materials evaluated for use on the probe.

2. As an uncoated probe material, tungsten is slightly superior to silicon carbide and silicon carbide is slightly superior to graphite.

3. The 90% W + 10% ZrO_2, flame-sprayed coating was the best coating tested for use on a molybdenum probe.

4. Self-bonded, silicon carbide coating was found to be the best of all the coatings tested for protecting a graphite probe.

5. The maximum surface temperatures on the models varied considerably from 1790°C (3255°F) to above 3200°C (5795°F) because of differences in the physical and thermodynamic properties of the various materials.

6. Further ballistic and mechanical tests should be conducted on the probe to establish definite specification requirements for the room and elevated temperature mechanical properties.

7. The effects of a slight amount of ablation, deformation, erosion or deposit on the nose tip of the probe on the efficiency of the probe should be investigated.

REFERENCES:

1. Stewart, J.D., "Estimated Fluxes on Proposed Pitot Tube for Testing in Water Arc", Gen Elec Co., Report No. AT117-91.

2. Dank, "Water Arc Tests on Nonmetallic Materials", Jour of Electrochemical Soc., Apr 1959, Vol 106, No. 4.

3. Molella, D.J., "Invest & Eval of Avail Probe Materials and Coatings for Use at 4000°F for 60 Sec", Picatinny Ars, Dec 1959, TM No. ME-5-59.

4. Smithells, Metals Reference Book, Interscience Publishers Inc., N.Y., 1955, Vol II, P.662.

5. Wilkes, "Evaluation of Coatings for Molybdenum", June 1961, Contract NOw 60-0321c.

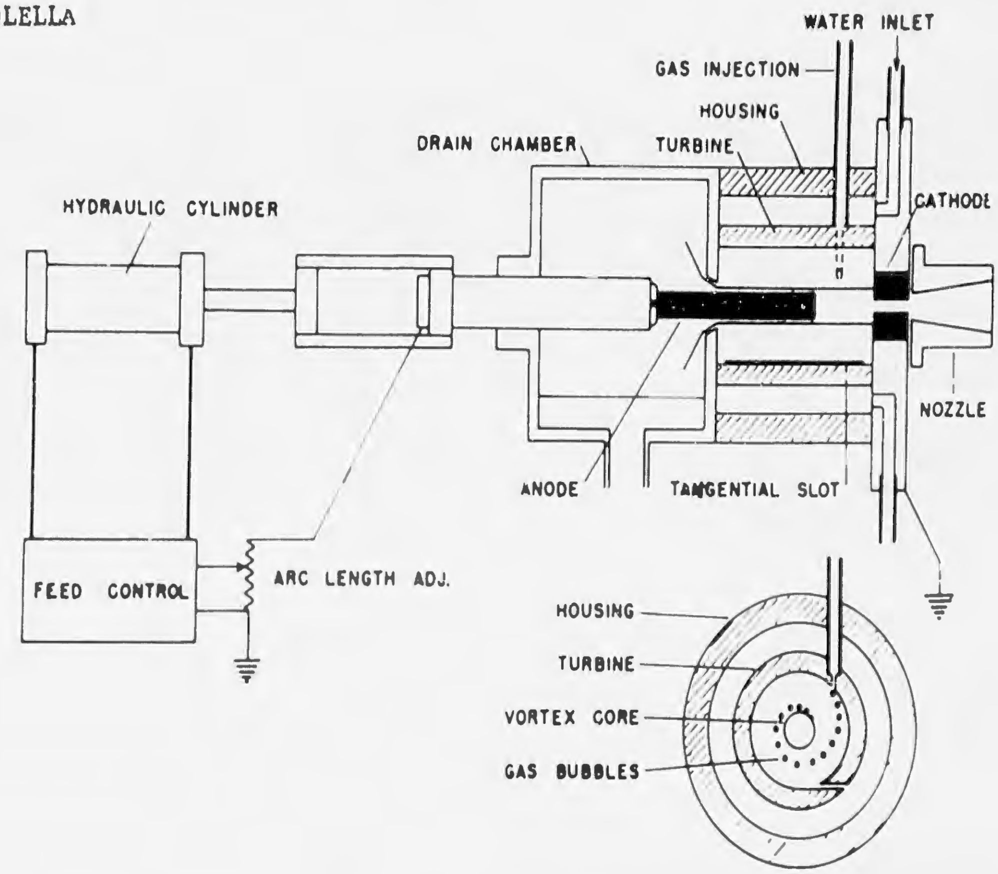


Figure I. Schematic diagram of water-stabilized electric arc

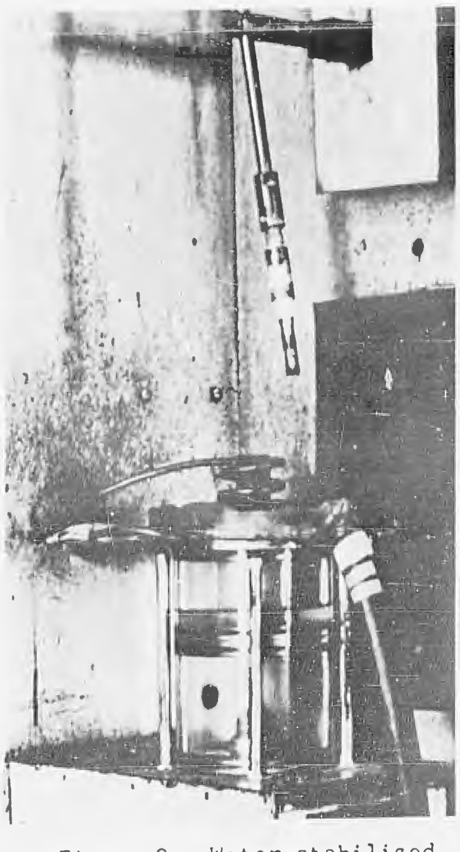


Figure 2. Water-stabilized electric arc apparatus before operation.

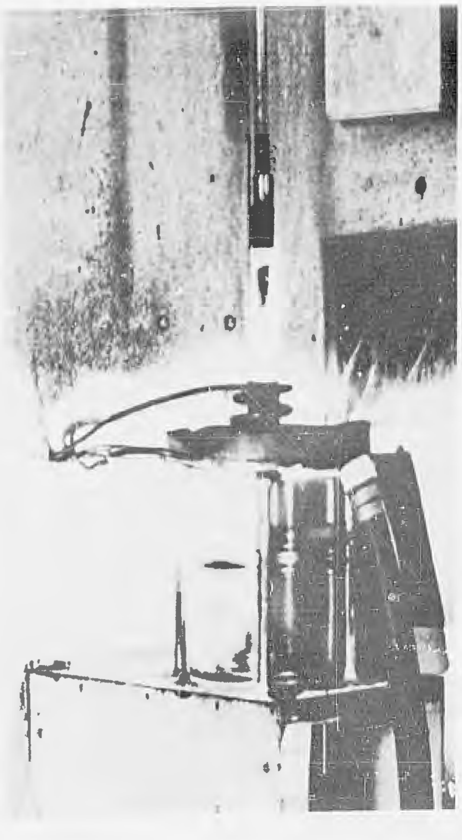
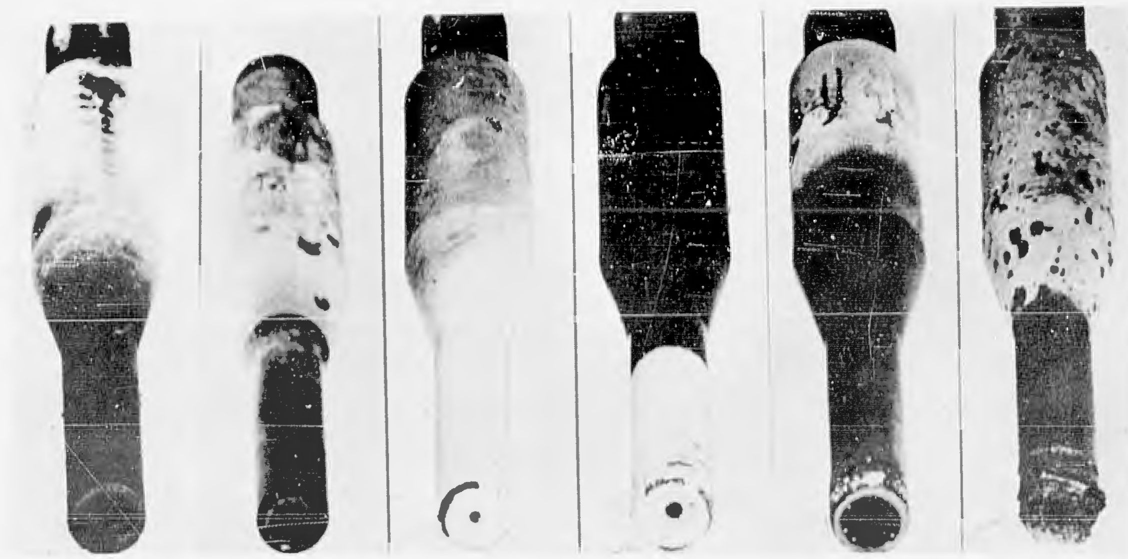
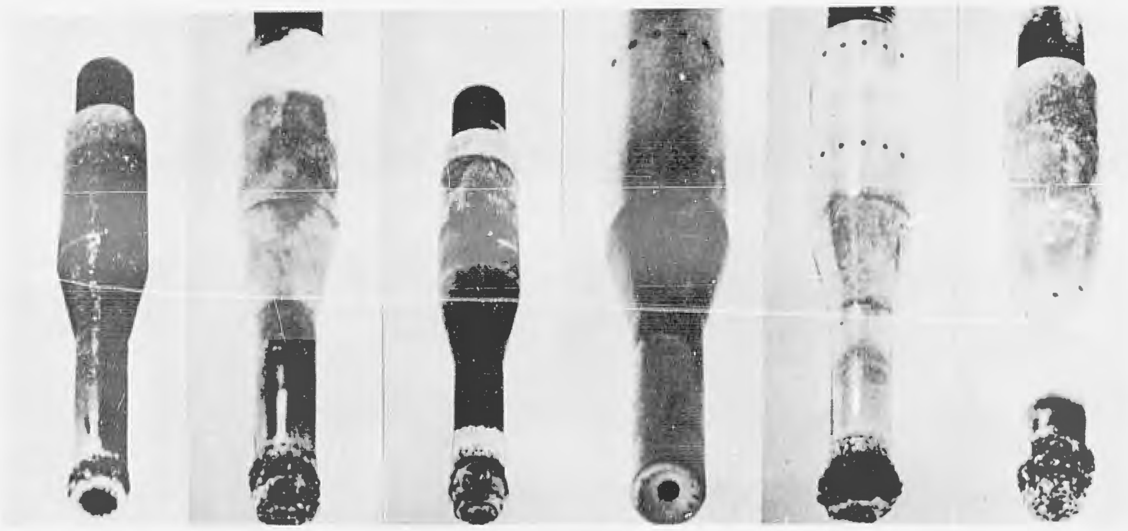


Figure 3. Electric arc during operation.



$90W+10ZrO_2$ Coating | Al-Cr-Si Coating | Gradated ZrO_2 Coating | ZrB_2 Coating | $MoSi_2$ Coating | $MoSi_2$ Coating



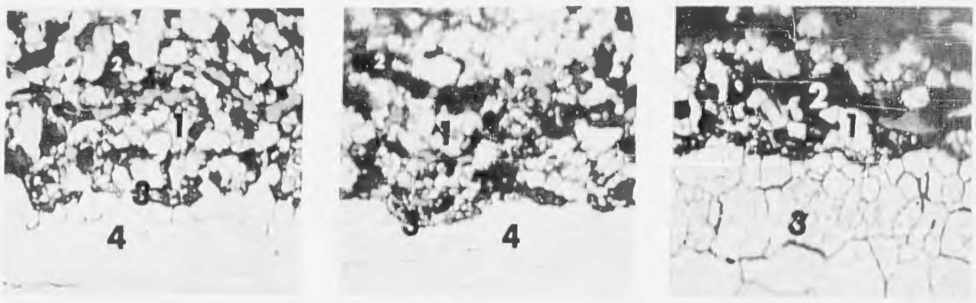
Durak-MG Coating | Durak-MG Coating | Al-Si Coating | W-2 Coating | W-2 Coating | LM-5 Coating

Figure 4. Photographs of molybdenum arc-tested probes with various protective coatings.

Graphite w/ SiC Coating | Graphite w/ AM.Lava Coating | Graphite w/ $90W+10ZrO_2$ Coating | Uncoated Tungsten | Uncoated SiC

Uncoated Graphite | Uncoated ZrB_2 | Uncoated Molybdenum | Uncoated WC w/Co Binder | Uncoated Cb-Ta-Zr Alloy | Uncoated Porous MgO w/Binder

Figure 5. Photographs of arc-tested graphite probes with various coatings (upper left) and photographs of various arc-tested, uncoated probes (upper right and bottom).



Before arc test. Note excellent bond at interface.
1. Coating 2. Porosity
3. Interface 4. Base metal

After arc test 0.50 inch from nose tip. Note slight damage to coating.
1. Coating 2. Oxidation
3. Interface 4. Base metal

After arc test at nose tip. Note no damage to coating.
1. Coating 2. Erosion and oxidation 3. Base metal

Figure 6 - Microstructure of Cermet (90% W + 10% ZrO_2) Coating on molybdenum
Etchant NaOH + K_3 Fe $(CN)_6$

Magnification - 1000X

Unetched 250X
Before arc test. Note excellent bond between 90 W + 10 ZrO_2 overcoat (1), tantalum undercoat (2) and base metal (3). Also note porosity (4) of base metal.

Unetched 500X
After arc test. Representative structure at and 0.50 inch from nose tip. Note complete oxidation of overcoat (1) and partial oxidation of remainder of overcoat and undercoat (2). Base material (3) remained intact.

Figure 7 - Structure of Cermet (90% W + 10% ZrO_2) coating on Graphite.

Before arc test. Note mechanical bond between coating (1) and base metal (4) and porosity (2) at interface (3)

After arc test 0.50 inch from nose tip. Note oxidation and melting of coating (1) and base metal (2)

After arc test at nose tip. Note catastrophic oxidation (1) of base metal (3) and floating of base metal to oxide (2).

Figure 8 - Microstructure of LM-5 coating on molybdenum
Etchant - NaOH + K_3Fe $(CN)_6$

Magnification - 1000X

DETERMINATION OF THE CRITICAL TORQUE INDUCING BUCKLING IN A TWISTED SPHERICAL SHELL SUBJECT TO INTERNAL OR EXTERNAL PRESSURE

C. C. MOW and M. A. SADOWSKY
WATERVLIET ARSENAL
WATERVLIET, NEW YORK

PROBLEM: The problem is illustrated by Figs. 2, 3 and 4. A spherical shell is twisted by two equilibrating torques applied by means of flanges at the poles of the sphere. The problem consists in determining the critical value (buckling value) of the torques and in analyzing the shape of the buckling impression (furrow) appearing near the flanges as buckling deformation. Internal pressure p may be present in the shell.

TOTAL BUCKLING VS. LOCAL BUCKLING: In case of a straight rod buckling under compression we have an example of total buckling. Every inch of the rod is deformed into a curved shape. There is no region of conspicuous concentration of deformation as against a region of relative absence of deformation. Every part, big or small, of the rod is engaged in deformation. In this sense, we call it a case of total buckling.

In case of spherical shells we may observe an entirely different picture: that of local buckling. By applying a negligibly small effort, a very thin spherical brass shell can be pushed in locally forming a combination of 2 spherical caps: a big cap of the undisturbed original sphere and a small cap in buckled position (Fig.1).

THE EXPERIMENT: LOCAL CHARACTER OF BUCKLING IN TORSION: Experiments made at the Watervliet Arsenal Laboratories have shown that the buckling of a spherical shell in torsion begins as local buckling concentrated close by the torque-transmitting flange of the torque producing machine. The shape of the buckled area is oblong-elliptic, stretching in an oblique direction from the edge of the flange. The buckled area resembles a fold or rather a furrow. Fig. 2 shows in a drawing the initial stage of buckling. Figs. 3 and 4 are photographs of more advanced stages in which the furrow is conspicuous enough to be seen and recognized in a photographic reproduction.

OUTLINE OF BUCKLING ANALYSIS BY THE ENERGY METHOD: Buckling equations are essentially energy balance equations. They state that the strain energy accumulated prior to the onset of buckling has reached a level sufficient to do the work necessary to produce the buckled shape. The external forces and moments pitch in and help in the performance of that work. The buckling equation of our problem will consist of an accurate balance of strain energy available for release through buckling and the work done by the external efforts on one side, and the energy required for infusion in the critical area to make it buckle, on the other side.

Each item mentioned above can be computed analytically. The strain energy of a shell consists of 2 parts: V_1, the strain energy due to the stretching of the middle surface, and V_2, the strain energy due to bending of the middle surface. The general principles for the computation of V_1 and V_2 have been established by A.E.H. Love and introduced to engineering analysis by S. P. Timoshenko. We will use methods of classical differential geometry on that occasion. The experiment shown in Fig. 8 will lead us to a reasonable assumption concerning the character of the local buckling process. From there on, we will arrive at the results by purely mathematical processes, making no further assumptions to suit the situation.

DIFFERENTIAL GEOMETRY OF THE BUCKLED SPHERICAL SHELL: We will use polar (spherical) coordinates r, θ, ϕ defined by

$$x = r\bar{p}\cos\phi$$

$$y = r\bar{p}\sin\phi$$

$$z = rp$$

in which

$$p = \cos\theta \quad , \quad \bar{p} = \sin\theta$$

Let a be the radius of the middle surface of the shell. A point P on the middle surface has polar coordinates

$$a, \theta, \phi .$$

Let the polar components of the displacement vector of P be

$$u_r = a\alpha$$
$$u_\theta = a\beta$$
$$u_\phi = a\gamma$$

in which

$$\alpha = \alpha(\theta,\phi),\quad \beta = \beta(\theta,\phi),\quad \gamma = \gamma(\theta,\phi)$$

are dimensionless functions of θ and ϕ. We will denote their derivatives with respect to θ and ϕ by subscripts, such as in

$$\frac{\partial\alpha}{\partial\theta} = \alpha_\theta \;,\quad \frac{\partial^2\beta}{\partial\theta\partial\phi} = \beta_{\theta\phi} \;, \;\cdots$$

THE BASIC STRESS FIELD: The stress field in the twisted shell at the critical value of the torque T prior to buckling is called the basic stress field. Its knowledge is important because in buckling a part of its strain energy is released for the formation of the buckled shape. The determination of the basic field, meaning the determination of all displacements and stresses in a spherical shell of radius a and thickness b under the action of the torque T_0 is a problem of 3-dimensional mathematical theory of elasticity. The solution, in polar coordinates r, θ, ϕ reads:

$$u_\phi = \frac{T_0 r}{4\pi a^2 h G}\left[\frac{\bar{p}}{2}\log\frac{1+p}{1-p} + \frac{p}{\bar{p}}\right]$$

$$\tau_{\theta\phi} = -\frac{T_0}{2\pi a^2 h}\cdot\frac{1}{\bar{p}^2}$$

$$u_r = u_\theta = \sigma_r = \sigma_\theta = \sigma_\phi = \tau_{r\theta} = \tau_{r\phi} = 0$$

We note that the only polar stress component that does not vanish is the shearing stress $\tau_{\theta\phi}$.

THE BUCKLING EQUATION: The buckling equation is an energy balance stating that the strain energy of the buckling deformation comes from two sources: from work done by external forces and from an influx of strain energy of the basic stress field. It can be found by systematically applying the mathematical theory of surfaces and space curves (differential geometry) to the occasion and it reads in the case when there is no internal pressure in the shell as follows:

$$\frac{T_0}{2\pi a^2}\iint \frac{1}{\bar{p}^4}\Big[(\alpha_\theta-\beta)(\alpha_\phi-\gamma\bar{p})\,\bar{p}+$$

$$+(\gamma p-\beta_\phi)(\beta p+\gamma_\phi+\alpha\bar{p})-\gamma_\theta(\alpha+\beta_\theta)\bar{p}^2\Big]\,dA=$$

$$=\frac{D}{4a^2}\iint\Big[(1+\nu)\Big(2\alpha+\alpha_{\theta\theta}+\frac{\alpha_\theta p}{\bar{p}}+\frac{\alpha_{\phi\phi}}{\bar{p}^2}\Big)^2+$$

$$+(1-\nu)\Big(\frac{\alpha_\theta p}{\bar{p}}-\alpha_{\theta\theta}+\frac{\alpha_{\phi\phi}}{\bar{p}^2}\Big)^2+4(1-\nu)\Big(\frac{\alpha_\phi p}{\bar{p}^2}-\frac{\alpha_{\theta\phi}}{\bar{p}}\Big)^2\Big]\,dA+$$

$$+\frac{Eh}{2(1-\nu^2)}\iint\Big[\Big(2\alpha+\beta_\theta+\frac{\beta p+\gamma_\phi}{\bar{p}}\Big)^2-$$

$$-2(1-\nu)(\alpha+\beta_\theta)\Big(\alpha+\frac{\beta p+\gamma_\phi}{\bar{p}}\Big)+$$

$$+\frac{1-\nu}{2}\Big(\gamma_\theta+\frac{\beta_\phi-\gamma p}{\bar{p}}\Big)^2\Big]\,dA$$

CHARACTERISTIC ASSUMPTIONS: The problem of solving the buckling equation consists of determining 3 functions $\alpha(\theta,\phi)$, $\beta(\theta,\phi)$, $\gamma(\theta,\phi)$ compatible with boundary conditions which will minimize the value of the torque T_0. Inspecting the structure of various integrals appearing we see that the bending energy (term with the coefficient D) depends on α alone. This points to a prominence of the function $\alpha(\theta,\phi)$ as compared with $\beta(\theta,\phi)$ and $\gamma(\theta,\phi)$. We now make our first characteristic assumption: we assume

$$\beta(\theta,\phi) \equiv 0 , \quad \gamma(\theta,\phi) \equiv 0 .$$

We drop from the bending energy integral all terms in α, α_θ and α_ϕ. This is the second characteristic assumption.

GEOMETRY OF THE BUCKLING FURROW ON A SPHERICAL SHELL: The basic guidance for judgment is given by the photographs Figs. 3 and 4 showing experiments made at the Watervliet Arsenal Laboratories with spherical shells buckling in torsion. Buckling begins as a localized effect concentrated close by the torque-transmitting flange of the torsion test machine. The shape of the buckled area is oblong-elliptic, stretching in an oblique direction from the edge of the flange. The buckled shape resembles a fold or rather a furrow. At the beginning, one single furrow appears. Photographic reproductions of that initial stage were unsatisfactory and have not been included in the present report. At a later stage of buckling, one or two more furrows of identical appearance would show up. The depth of the furrows at that stage was sufficient to produce illumination effects giving good photographic reproductions. The photographs included here show a stage at which 2 furrows have been formed. The theoretical analysis refers to the initial stage of the formation of the first furrow.

The buckling equation gives

$$T_0 = 8.583 \frac{E h^2 r^2}{(1-\nu)\sqrt{1+\nu}\, a}$$

in which r is the radius of the flange bonded to the shell. For the minor semiaxis of the elliptic furrow we have

$$b = \frac{1.354}{\sqrt[4]{1+\nu}} \sqrt{a h}$$

THE COMPLETE BUCKLING DIAGRAM IN PRESENCE OF INTERNAL PRESSURE p: Let T_0 be the buckling torque in absence of pressure, and p_0 the external buckling pressure in absence of a torque. The value of p_0 resultant from the present computation is

$$p_0 = \frac{(\sqrt{6}+3\sqrt{2})}{3(1-\nu)\sqrt{1+\nu}} \cdot \frac{Eh^2}{a^2}$$

The complete buckling diagram is shown in Fig. 5. It extends over the ranges of dimensionless variables

$$-1 \leq \frac{p}{p_0} < \infty \quad , \quad 0 \leq \frac{T}{T_0} < \infty$$

in the range

$$-\sqrt{\frac{2}{3}} \leq \frac{p}{p_c} < \infty$$

the diagram is a straight line given by

$$\frac{T}{T_0} = 1 + \sqrt{\frac{2}{3}} \cdot \frac{p}{p_0}$$

in the range

$$-1 \leq \frac{p}{p_0} \leq -\sqrt{\frac{2}{3}}$$

the diagram is part of the ellipse

$$\frac{T}{T_0} = \frac{1}{\sqrt{3}} \cdot \sqrt{1 - \frac{p^2}{p_c^2}}$$

According to this result, it is possible to make a membrane shell torque resistant by inflating it with internal pressure. The rectilinear part of the diagram in Fig. 5 becomes

$$T = \pi a r^2 p \qquad p \geq 0$$

for the membrane shell.

CONCLUSIONS: Based on exact methods of differential geometry and on accepted engineering approximations in applying those methods to shell theory, it has been possible to successfully deal with a pressurized spherical shell buckling under action of twisting torques. The critical relation between the values of the torque and pressure under whose simultaneous action buckling takes place has been obtained. Inflation of the shell by internal pressure increases the resistance against torsional buckling.

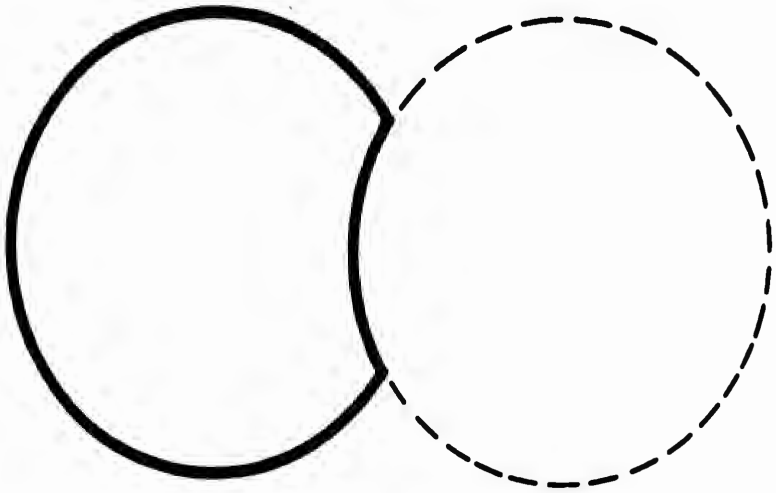


Figure 1. Local buckling of a spherical shell by formation of a reflected cap.

Figure 2. Buckling of a spherical shell under twisting torques.

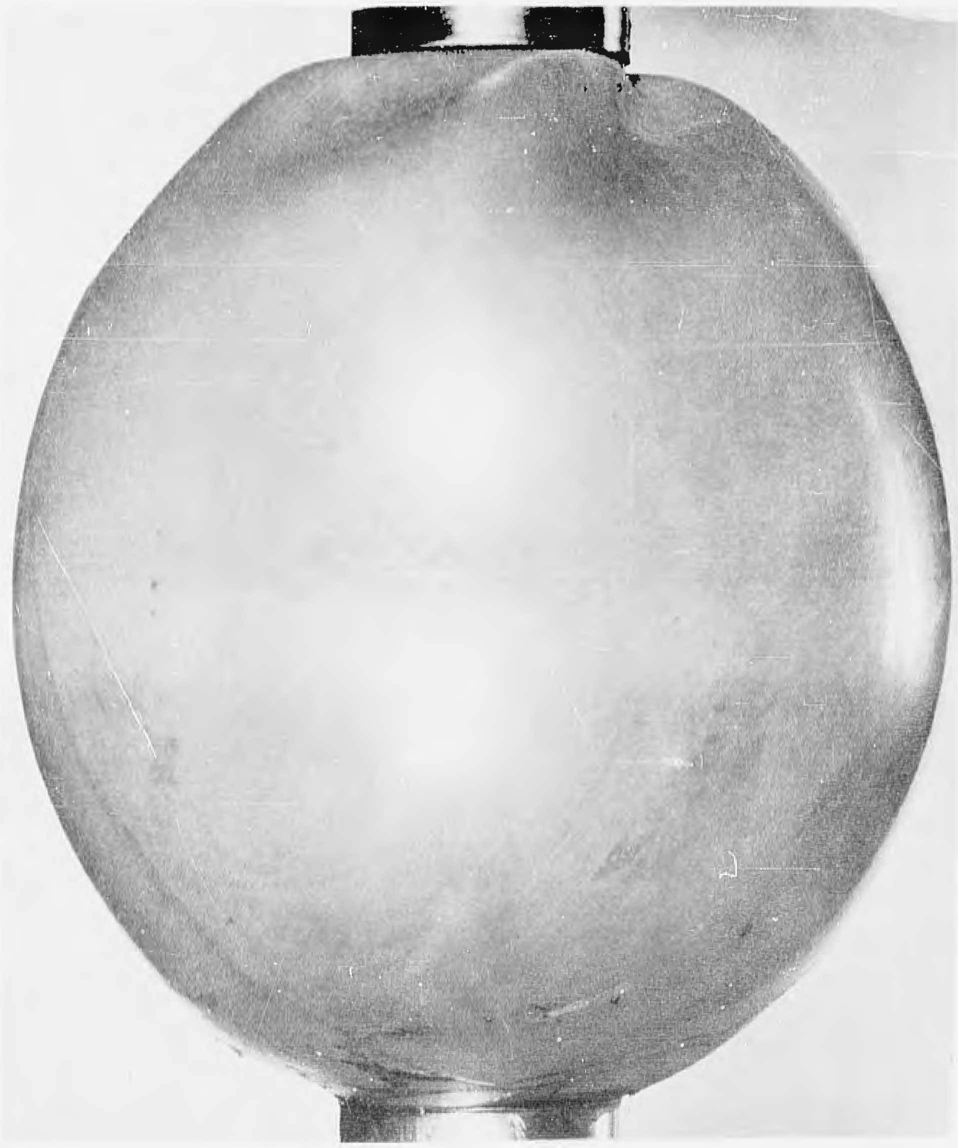


Figure 3. Actual test photograph of a spherical shell in local buckling under twisting torques.

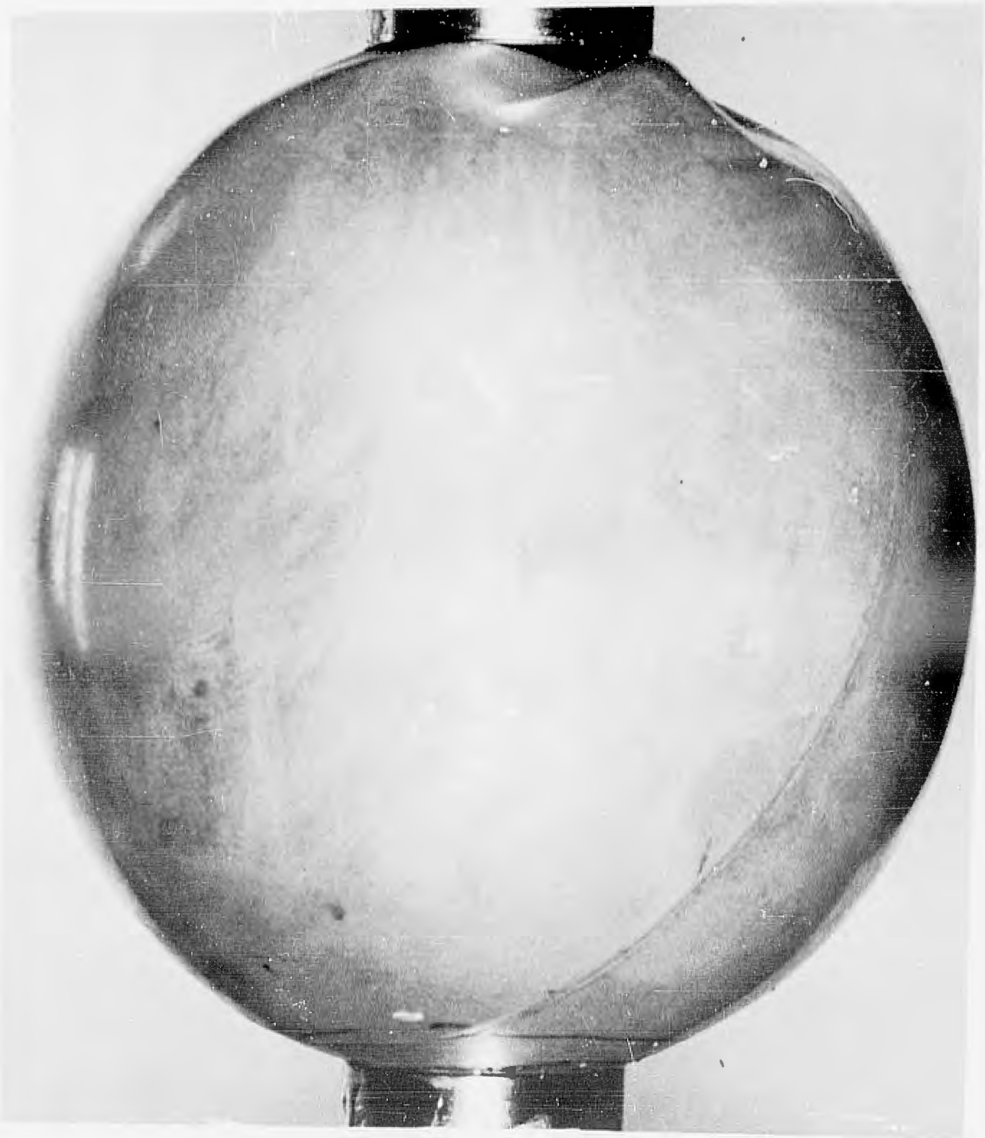


Figure 4. Actual test photograph of a spherical shell in local buckling under twisting torques.

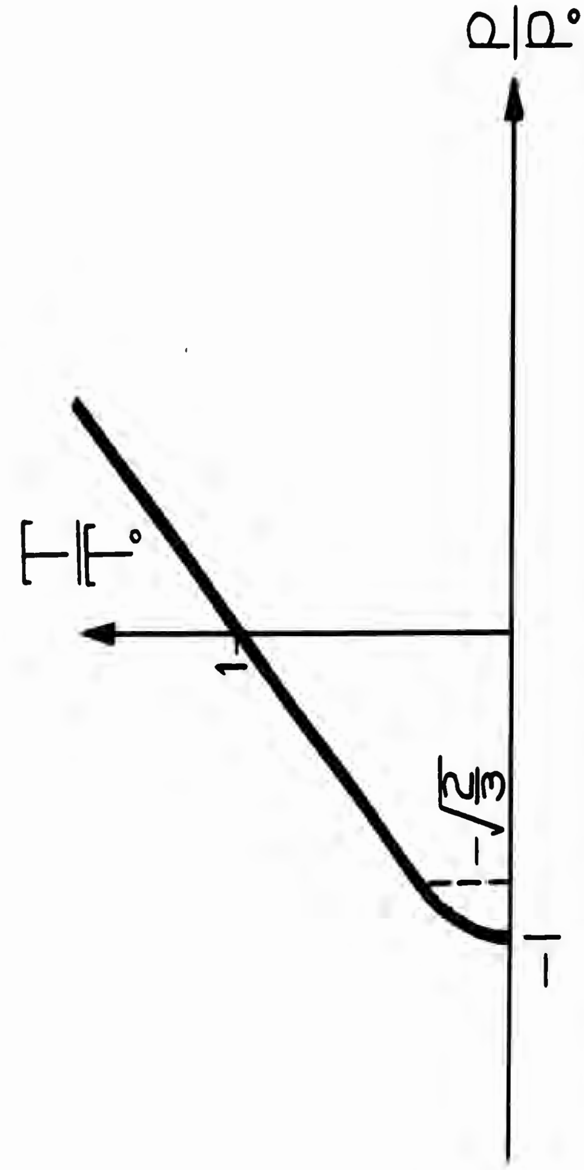


Figure 5. The complete critical diagram

PROCESSES OF PHOTON AND EXO-ELECTRON EMISSION FROM ALKALI AZIDES

H. J. MUELLER, H. D. TILLER, and G. D. SINGER
BASIC RESEARCH GROUP
USAERDL, FORT BELVOIR, VIRGINIA

INTRODUCTION

Inorganic azides are metastable compounds of the type XN_3; they decompose with varying ease when subjected to energy interaction (irradiation, heat). Once initiated, in some azides the decomposition becomes self sustained and develops into an explosion (PbN_6, CuN_6), whereas substances like KN_3, NaN_3, BaN_6, for instance, do not detonate, at least under normal conditions. It is this wide range of metastability found among the azides which makes their investigation interesting from the military point of view. Particularly the mechanism of the decomposition and its parameters deserve attention in the light of modern solid state physics since it has been recognized that the classical hot spot model of initiation, which only takes in consideration energy interaction with a perfect solid and in the form of heat, is over simplified; in fact, structure defects, photons, and free electrons also play an important role for the decomposition process and its self sustained propagation through the lattice. Hence, at the present time intensive research is going on with the objective to reveal the nature of metastability, employing advanced techniques of experimentation such as ESR, IR-spectroscopy, mass spectrometry, tracer methods, etc. Part of these efforts is the study of photon and electron emission phenomena associated with the decomposition of such typical and, from the structural point of view, simple metastable compounds like the inorganic azides. To start with we chose the azides of potassium and sodium (KN_3, NaN_3) because there is already available a good deal of data concerning their crystallographic, electronic, and defect structure, which fact facilitates considerably the interpretation of new results.

The occurrence of electron emission from solids under conditions which neither permit photoemission nor thermionic emission is known under

the name exo-electron emission.[1] Many solids emit electrons subsequent to irradiation on warming up, during and after plastic deformation, and during phase transition. The emission currents are usually very small (to the order of 10^{-15} A/cm^2 or less) and their intensity decreases with progressing time after the exciting event (irradiation, deformation, etc.). Although there are many open questions with respect to the mechanism of the exo-electron emission, this phenomenon is fairly well understood in the case of typical ionic crystals, for example, NaCl, KCl, CaF_2, etc. Here the exo-electron emission appears to be closely related to luminescence and to processes involving color centers. Electrons trapped at suitable lattice defects are lifted by lattice vibrations or photon interaction into the conduction band of the crystal from where they either may recombine with an electron-deficient defect whereby generally luminescence occurs or they may leave the crystal lattice entirely as exo-electrons by virtue of a mechanism not yet known in detail. This model is essentially based on the observation that in pure alkali halides temperature dependency of luminescence, exo-electron emission, and electric conductivity exhibit parallel course as illustrated in Figure 1. It was of basic interest to investigate this correlation also in the case of alkali azides which behave in many regards similar to the respective alkali halides.

EXPERIMENTAL

Samples of pure KN_3 or NaN_3 either in the form of crystals or as a crystalline powder have been investigated by employing the glow curve method, i.e., the particular physical quantity (light emission, electron emission) was measured as a function of the steadily rising temperature of the sample. The glow curves were taken between -196°C and +400°C. The luminescence was detected by a photomultiplier with a quartz window (RCA C-7267) responding between 7000 and 2000 Å. The electron emission was measured by an open Geiger-Müller counter as described in the literature reference cited in the preceding footnote. Irradiation of the sample with UV light was performed in vacuo at the temperature of liquid nitrogen. A mercury arc was used as light source; the infrared component of the radiation was filtered by a layer of 20 mm water.

RESULTS AND INTERPRETATION

The results obtained are shown in Figures 2 and 3. For these diagrams a set of typical glow curves has been selected.[2] The height of the

[1] An up-to-date review of this topic was recently published by one of us (H. J. Mueller, "Exo-electron Emission and Related Electron Emissions," U. S. Army Engineer Research & Development Laboratories, Research Report 1704-RR, 20 December 1961).

[2] The data on the luminescence of NaN_3 were obtained by Mr. H. D. Tiller while working on his Master's thesis at Clemson College, South Carolina.

glow curve maxima fluctuated from sample to sample, obviously depending on the degree of imperfection of the specimen. The position of the maxima, however, was reproducible with good accuracy (±5°C).

a. Potassium Azide

Non-irradiated samples of KN_3 showed neither luminescence nor electron emission between -196°C and 200°C. A broad luminescence peak was observed at about 280°C, a peak of the electron emission occurred at 330°C. At 350°C the samples melted without causing noteworthy emission effects. Samples irradiated with UV-light at -196°C gave glow curves rich of details. The luminescence glow curve exhibited a group of four well resolved maxima in the low temperature range (-165, -130, -80 and -40°C). The electron glow curve, however, showed only a small emission maximum corresponding to the largest luminescence peak at -130°C. Above 0°C both luminescence and electron emission passes through several maxima. Noteworthy is the appearance of corresponding peaks at 210/215°C, both in the luminescence and electron glow curve. The 280°-maximum of the luminescence and the 330°-maximum of the electron emission are obviously identical with those observed at the same temperatures in the glow curves of non-irradiated samples. The spectral characteristics of the emitted luminescence light was examined by interposing optical filters between sample and photomultiplier since the low intensity of the luminescence, which was invisible even for the dark-adapted human eye, did not permit the use of a spectrometer. It was found that light was essentially emitted in the wave length range between 4000 and 6000 Å. It has to be emphasized that no light emission in the near ultraviolet was observed, although the employed photomultiplier was sensitive down to the quartz cut-off. The largest exo-electron emission observed was of the order of $5 \cdot 10^5$ electrons/cm^2.

If one compares these results with those of similar experiments on alkali halides one notes that the good correspondence between the maxima of the luminescence and the exo-electron emission, which is considered to be typical for substances with ionic bond, does not exist in the case of KN_3. Moreover, even non-excited fresh KN_3 samples show both luminescence and exo-electron emission (at 280 and 330°C), which has never been observed on fresh and non-excited alkali halides. In order to understand the details of the glow curves we shall attempt to correlate the observed emission peaks to electron processes known to occur in UV-excited KN_3, to which luminescence and exo-electron emission are closely related. The picture we have of these processes was essentially obtained on the basis of optical and electrical measurements (1).

On irradiation by UV-light at the temperature of the liquid nitrogen the originally colorless KN_3 becomes violet. This coloration is attributed to the formation of F-centers (i.e., one electron located in a negative ion vacancy). The irradiation also creates V-centers (an anion stripped of one electron trapped at a positive ion vacancy) as well as R and M centers (F-centers associated with one or more vacancies). Whereas the F-centers are reported to be stable at the temperature of dry-ice, the V-centers disappear at fairly low temperatures on warming up of the colored crystal. Hence, we believe that the luminescence with peak at -165 and -130°C is caused by the annihilation of V-centers. In KN_3 this is possible in two ways, at least. Like in alkali halides, the V-center may dissociate thermally in vacancies and holes; the latter are able to move even at low temperatures through the lattice whereby they may recombine with electrons trapped in F-type centers. However, the mobile holes may also associate with each other under formation of molecular nitrogen. In both processes the excessive energy of the originally excited recombination complex will be dissipated, at least partly, in the form of photons and, thus, luminescence can occur. Neither of these processes requires electron transitions via the conduction band, the occurrence of which is consid[illegible] to be essential for the exo-electron emission from ion crystals. Hence, it is not surprising that practically no electron emission was observed in this temperature range. The weak electron emission between -100 and -140°C is probably caused by absorption of luminescence photons in the lattice; the wave length of the luminescence light is certainly short enough (4000 - 6000 Å) to cause external photo effect, the threshold of which we found to be 4500 Å for UV-irradiated KN_3 at room temperature.

Between -78° and room temperature the violet coloration changes to blue. This is attributed to the transition of F-centers into R-type centers by simple association of the former with other structure defects. As one should expect, exo-electron emission does not occur during this process either. In this temperature interval, however, a group of two luminescence maxima of lesser intensity are observed (-80°C and -40°C). We propose to ascribe at least the maximum at -40°C if not both, to the formation of R-centers.

Between 70° and 100°C the R-centers disappear due to the dissociation of the F-complexes into vacancies and electrons, and metallic potassium forms in the lattice. Applying the theory of glow curves we calculated that the process responsible for the electron emission at 100°C should have a thermal energy of activation of 0.8 eV, which is in KN_3 equivalent to an optical energy of 2,43 eV. This, however, is roughly the energy required to promote F-center electrons to the conduction band and thus it seems logical to associate both electron emission and luminescence

observed in the respective temperature range with the dissociation of F-centers.

The large electron emission peak at 215°C which also corresponds to a well resolved maximum in the luminescence glow curve cannot be interpreted satisfactorily on the basis of our present knowledge of electron processes in KN_3. We were informed, however,[3] that in this temperature range, in which the crystals exhibit a yellow color, a new, sharp optical adsorption band at approximately 9300 Å develops and an ESR signal is observed which seems to be due to the presence of free radicals. Since it is obvious that any process in which such electronically excited complexes participate will lead to the emission of photons and/or electrons we propose to attribute tentatively the photon and electron emission peaks observed at 210° and 215°C, respectively to a process involving these hypothetical free radicals.

Both the 280°-luminescence bump and the 330°-electron emission peak are different from the previously discussed maxima inasmuch as they occur also in the glow curves of the non-irradiated material. This indicates that the emission is rather to ascribe to pure thermal causes than to radiation-induced effects. In this temperature range, which is very close to the melting point of KN_3, measurable evolution of gaseous nitrogen indicates the occurrence of thermolytical decomposition and the mobility of defects and ions in the lattice is highly increased, the latter fact being demonstrated by the coagulation of metallic potassium to specks of colloidal size above 200°C. Thus we believe that the observed luminescence and electron emission is associated with chemical reactions of decomposition species with each other or with the surrounding gas atmosphere. This interpretation is supported by experiments in which the luminescence at elevated temperatures was observed while the sample was kept in various gas atmospheres (He, A, N_2, air) as well as in vacuum; the luminescence above 200°C was definitely increased by the presence of air.

The fact that the 330°-exo-electron maximum is usually several times higher for UV-irradiated samples than for non-irradiated material can be interpreted as a result of the photolytical decomposition during the initial UV-irradiation which leads to the formation of nucleii of metallic potassium in the crystal. The presence of metallic potassium, however, accelerates catalytically the thermal decomposition and this fact might account for the stronger photon and electron emission above 250°C in pre-irradiated samples.

---

[3] Private communication by Mr. G. J. King and co-workers, Basic Research Group, USAERDL.

For irradiated KN_3 a sharp luminescence peak characterizes the break down of the lattice at the melting point. Since only a slight indication of this peak is found in the glow curves of non-irradiated samples its appearance is rather due to a chemical process - for example, the reaction between colloidal potassium metal, the end product of the radiation induced color center processes in KN_3, and residual oxygen - than to triboluminescence associated with the sudden release of thermally induced strains in the collapsing lattice. Beyond 350° both measurements of the photon and electron emission became unreliable since evaporating potassium azide and potassium metal condensed on the window of the photomultiplier as well as in the interior of the Geiger-counter.

b. Sodium Azide

In NaN_3 the situation was found to be considerably less complex than in KN_3. Fresh, non-irradiated samples of NaN_3 showed neither photon nor exo-electron emission. UV-irradiated NaN_3, however, exhibited strong luminescence in the low temperature range although no simultaneous exo-electron emission was observed. On the basis of optical and ESR measurements (1)(2) it is believed that on irradiation of NaN_3 with UV-light at the temperature of liquid nitrogen F_+^2-centers (i.e., two vacancies plus one electron) and, to a lesser extent, F-centers are formed. The latter are unstable even at -196°C and they bleach readily on warming up, whereas the F_+^2-centers still exist at -80°C; they disappear, however, on further warming to room temperature. By contrast to KN_3, no V-centers are formed by UV-light in NaN_3. Hence, we attribute the luminescence observed between -196 and -80°C to the thermal bleaching of F-centers. The -60°C peak has to be associated with the dissociation of the F_+^2 centers. At these temperatures, however, thermal bleaching of both F and F_+^2 centers is presumably a mobility effect without electron transitions via the conduction band; for example, the transfer of the F-center electron to an adjacent metal ion so as to give metallic sodium. Thus, it is understandable that no exo-electron emission occurs simultaneously to the observed luminescence. Above 200°C NaN_3 starts to decompose rapidly and the increasing emission of electrons is believed to be due to this fact.

The luminescence peak at 100°C and the small peak in the electron glow curve at 330°C are not always observed. Their erratic appearance suggests that they are due to contaminations on the surface of the NaN_3 crystals. There is electron microscopical evidence[4)] that even fresh NaN_3 reacts easily on short exposure to air, carbonate presumably being the end product.

4) Private communication by Dr. J. Joebstl, Basic Research Group, USAERDL.

ACKNOWLEDGMENT

The authors acknowledge gratefully the valuable support of this work by their colleagues of the Basic Research Group. The interest of Dr. Z. V. Harvalik, Group Director, is much appreciated.

REFERENCES

(1) B. L. Evans, A. D. Yoffe, and P. Gray, Chem. Rev. 59, 515 1959.
F. C. Tompkins F.R.S., and D. A. Young, Proc. Roy. Soc. A236, 10 1956.
J. Cunningham and F. C. Tompkins F.R.S., Proc. Roy. Soc. A251, 27 1959.
H. G. Heal and J. P. S. Pringle, J. Phys. Chem. Solids, 15, 261 1960.
S. K. Deb, J. Chem. Phys. 35, 2122 1961.

(2) G. J. King, B. S. Miller, F. F. Carlson, and R. C. McMillan, J. Chem. Phys. 35, 1442 1961.

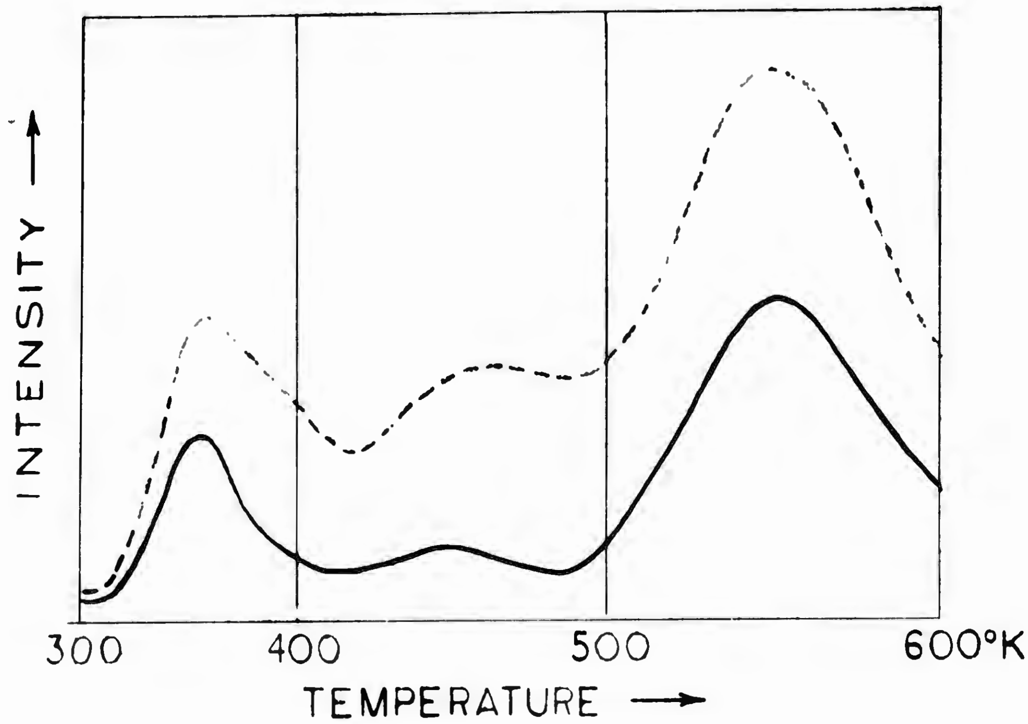


Figure 1: Thermoluminescence (——) and exo-electron emission (----) from x irradiated CaF_2 (after F. Bohun).

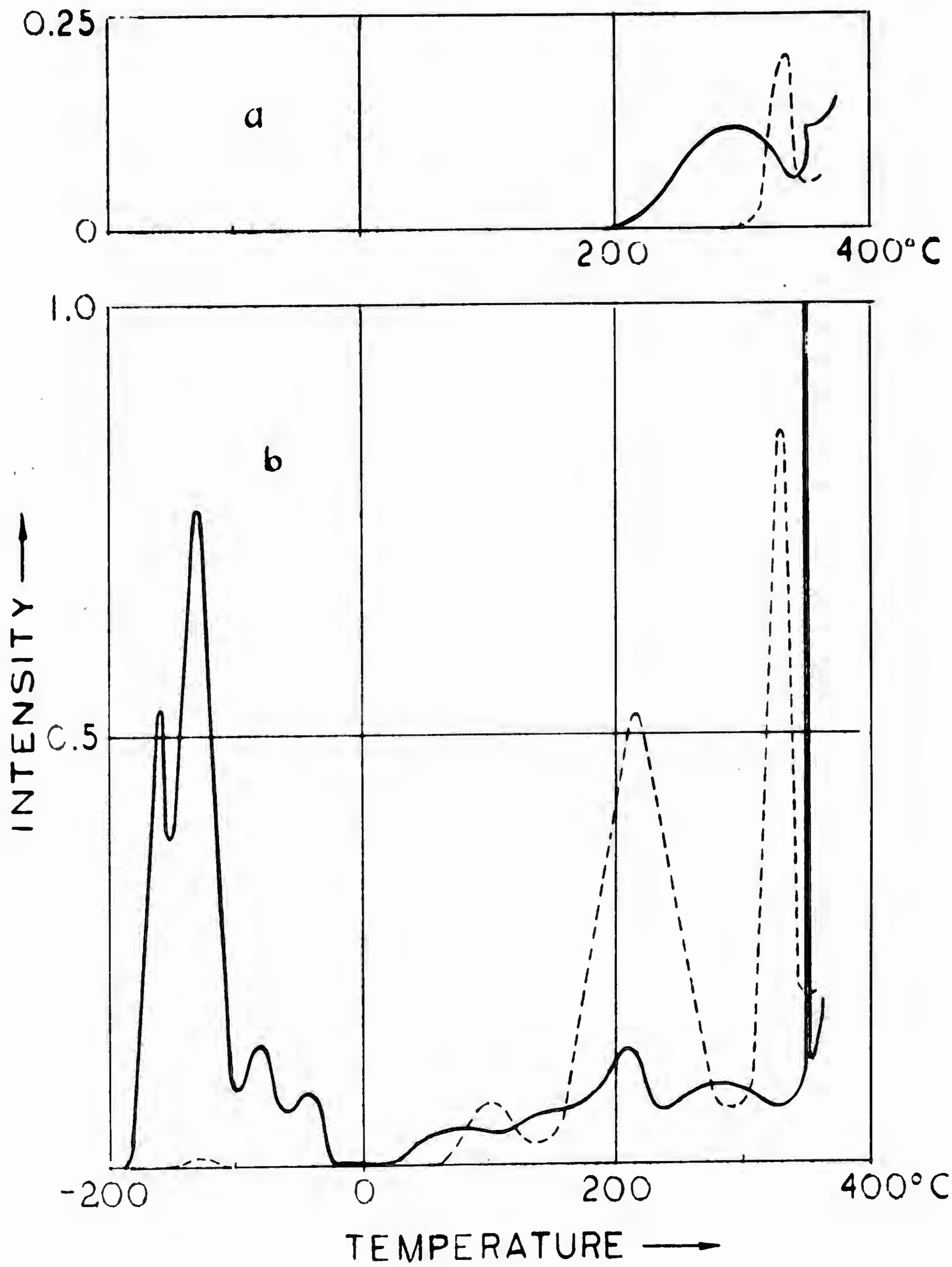


Figure 2: Thermoluminescence (——) and exo-electron emission (----) from potassium azide. a) non-irradiated b) UV-irradiated

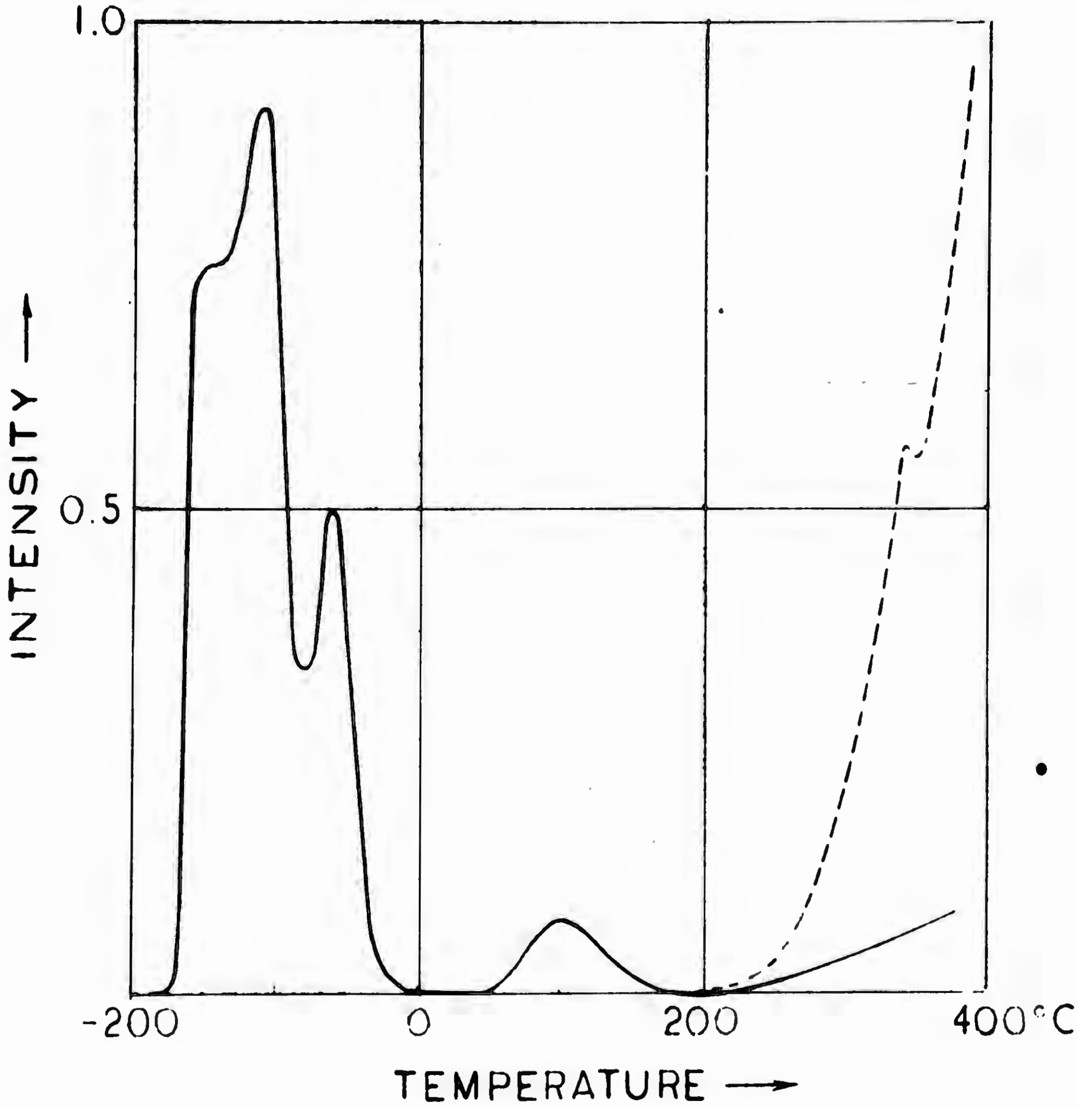


Figure 3: Thermoluminescence (——) and exo-electron emission (----) from UV-irradiated sodium azide.

FIELD EXPERIMENTATION AS AN AID TO OPERATIONS RESEARCH

F.J.MURDOCH,JR,COL, GS, and CARL C.EDMONDSON, COLONEL, GS *
US Army Combat Development Experimentation Center
Fort Ord, California

Just as science played its first part in warfare in the development of the materials of war, so the methods of operations research were first applied by the military to the solution of equipment problems. Over the last decade, science and its methods have made an over-expanding contribution -- not only to the development of the tools of war but also in their application. Today warfare has become quite literally a mixture of art and science, with the soldier and the scientist joined in common efforts to solve our problems of national defense ...

Operations research is generally concerned with the application of the principles, methods, and tools of science to the study and analysis of the efficiency and functioning of varied and complex organizations of men and machines. If one accepts this definition, then we have a common starting point for both the soldier and the scientist. It was during World War II and the Korean War that operations research made its greatest contributions in the military field. These contributions to ground, naval, and air warfare are well known and will not be reiterated here. Since then, operations research has not only continued in equipments and weapons systems, but has encompassed even larger and more complex problems.

The approach has been largely mathematical and analytical, based on models and theory. But while the scientist has his methodology, his techniques and technical tools, these alone are not enough. If realistic and practical solutions are to be found, the scientist must base his theories and build his models on significant, accurate, and valid information. In time of peace, the real experiences of actual operations are hard to reconstruct. Consultations with experienced military

personnel are often valuable. However, their experience inevitably reflects the past, and, even when projected to the future with imagination and foresight, has to be put to the test. Only through field experimentation can these concepts for the future be realized and evaluated.

It was for this purpose that the US Army in 1956 established the Combat Development Experimentation Center (CDEC) to prepare, conduct, and evaluate, with maximum objectivity and scientific control, experiments with concepts, organizations, tactics, doctrines and procedures for future combat.

To carry out this program, CDEC is provided with military and scientific personnel for planning, supervising, and executing field experimentation, and analyzing and reporting results. A large area of some 250,000 acres, and troops and equipment are available for tactical operations. This field laboratory provides for the execution of experimental concepts by men and machines under realistically simulated combat conditions. These exercises produce both quantitative and qualitative data under conditions that reflect operational degradations, human behavior, functioning of machines, enemy measures and counter-measures, and other elements that influence the systems under evaluation. Instrumentation and simulation means have been developed and the methodology of field experimentation is continually being improved.

To date the program for field experimentation has been concerned with evaluating the tactical application of new concepts of equipments, organizations, weapons systems and surveillance and target acquisition systems. CDEC is not an equipment-testing agency. However, we do measure the performance of equipment in a simulated war. By simulation, concepts for new equipments are evaluated in terms of their tactical application. Thus, the need for, and capabilities of, proposed equipments may be determined prior to expending time and money on the development of an item.

The advent of the tactical nuclear weapons to the battlefield has stimulated new concepts of organization based on increased dispersion, surveillance, mobility, and more critical command and control. Experimentation has been conducted on the controllability, mobility, target-acquisition ability, vulnerability, sustainability, and the destructive force of new organizations. Experimentation has also been conducted on the vulnerability of low flying aircraft to ground fires including the REDEYE infrared homing missile. In addition, many data from logistical tests have been collected on the problems of supply and resupply, maintenance,

and medical evacuation. Not only has all of this experimentation produced specific conclusions and recommendations within itself, but it has also provided valid and realistic data for general use in war gaming and operations research activities of both CDEC and other organizations.

This valuable support for war gaming and other operations research studies is accomplished in two ways. First, field experimentation provides valid and realistic data for direct input into existing war game models. Second, it provides a firm reference point for the logical development and refinement of models, both computer simulation models and terrain board models, that are specifically constructed to expand and clarify the questions and solutions produced in the field environment. Since both field experimentation and operations research models have different inherent limitations, it is to be expected that the proper interdependent use of these supplementary devices at CDEC can take us much further in answering a research question than the use of either one (or both) independently.

Although CDEC is relatively young, being the US Army's first attempt in field experimentation, it has and can make major contributions to other research activities. One example is a US Army project with the University of Michigan. Among the objectives of PROJECT MICHIGAN is the development of techniques and procedures for assigning missions to various surveillance devices and for correlating the returns from these devices. This has involved laboratory experiments using a fully manned and equipped surveillance data processing center operating on simulated information. Field data from CDEC are used to construct and validate the simulation models, both of the basic battlefield situation and of the surveillance devices returns. Without this realistic field data no trust could be put in the simulation. Field experimentation can also provide knowledge as to the type and quantity of information that a commander requires to accomplish his mission, as opposed to a requirement to develop a device that sees all and knows all.

The Operations Research Office of John Hopkins University, the predecessor to the Research Analysis Corporation, (RAC), not only has used data from CDEC experimentation but has actively participated with CDEC in certain field experimentation. For example, in the aircraft vulnerability experiment, ORO gathered data on the first three of the following objectives of the experimentation.

a. To determine the vulnerability of low-flying aircraft to ground fires of forward area weapons anticipated for the 1965 period (ORO and CDEC).

b. To determine the effects of mask angle on reaction time by ground troops in detecting and firing on aircraft (ORO).

c. To determine the effect of aircraft speed up to Mach .8 on the vulnerability of low flying aircraft subjected to forward area ground fire (ORO and CDEC).

d. To determine the effects of sequential penetrations by several aircraft on performance of troops equipped with forward area weapons (CDEC).

e. To determine the effects on the performance of forward area troops when required to make positive identification of aircraft as friendly or enemy (CDEC).

f. To determine the effects that various altitudes and directions of flight of drone aircraft over forward area troop dispositions will have on the ability of those troops to acquire and track such targets (CDEC).

There is a major category of research activity to which the data from field experimentation is of great potential value--battle simulation models. Field experimentation can provide a basis for determining model details, parameters, and input data. The following are types of information available from field experimentation:

a. Terrain and other features of environment.

b. Details of organizations and equipment.

c. Type of actions and tactics employed.

d. Roles of weapons systems.

e. Target sighting, frequency and selection priorities.

f. Rates of fire, firing delays and ammunition requirements.

g. Detailed records of:

(1) Hit and kill assessments, showing time, location and units involved.

(2) Time units are neutralized or suppressed by enemy fire.

(3) Use of non-organic fire support atomic weapons, CBR, etc.

(4) Record of all decisions and orders by commanders.

(5) Records of equipment and communication failures.

The provision of this type of information to the research analyst gives him a thorough knowledge of the elements that go into the system and allows him to apply realism, constraints and practicality to his theoretical analysis.

It must be admitted, however, that the relationship between field experimentation and operations research is a two-way or mutually interactive one. Operational gaming and computer simulation can provide "inputs" for field experimentation as well as vice versa. By definition of the system to be examined, by delineation of parameters, and by specification of relevant variables and hypotheses, the focus of these theoretical studies can help to sharpen actual field experimentation. This has obvious consequences in terms of economy of effort and maximization of results. A small example of this two-way tie between field experimentation and operations research is provided by an examination recently made at CDEC of the adequacy of a casualty evacuation system. Given the experimentation assumptions and limitations and estimates of certain relevant factors, computer simulation of the evacuation system established the casualty loads and conditions under which the system should be tested in the field. In turn, the results of the field experiment provided actual distributions as inputs in place of estimates, and served to test the adequacy of the simulation model.

The Aircraft Survivability experiment recently conducted by CDEC will illustrate how a field experiment can provide a sound basis for computer simulation. The experiment objective was to determine the survivability of low-flying aircraft opposed by a REDEYE weapon system. More specifically, it was necessary to determine:

a. The percentage of all flights which will be detected.

b. The percentage of detected flights which will be engaged.

c. The percentage of planes on detected flights which will be killed.

d. The extent to which the weapon system can be saturated by multiple plane and simultaneous flights.

The field experiment involved two opposing battalion-size forces engaged in normal tactical operations. A REDEYE system operated as an organic part of each battalion and a reasonable flight schedule of typical missions was flown for each battalion. The types of field data collected for each aircraft flight pattern included:

a. Actual flight paths.

b. Actual weapon locations.

c. An actual "in view" pattern for each weapon (both in terms of where plane was and duration of in-view periods).

d. Gunner performance characteristics (e.g., time required to make the friend-or-foe identification, time of lock-on, time of fire, etc.)

These data were then used as the basis for a computer simulation study of the system saturation problem and of the effect of varying factors such as the single-shot kill probability and firing doctrine. Thus the field experiment provided real input data on target availability for the computer simulation. Of obvious necessity, however, these data from the field could not take "kills" into account; we could not actually shoot down planes, especially since some of them were Navy jets! The computer simulation, however, provided the vehicle for extending the field data the remaining step to obtain information on aircraft survivability. We see, then, how the inherent limitations of field experimentation and computer simulation are complementary so that the integrated use of the two techniques provides information of greater quantity and validity than could be obtained by the independent use of field experimentation and simulation.

At the other end of the spectrum, field experimentation can be used to test the validity of mathematical and analytic solutions. Detailed experimentation verification of results of model simulation will do much to reassure the faith of those who feel that too much reliance is placed on the mathematician with his mathematical methods, models and theory. At the same time, field experimentation gives the operations researcher an opportunity to visualize, experience and examine future systems and thus reduce his need to rely on the scientific crystal ball.

In summary, this brief survey has shown that field experimentation can:

a. Provide simulated battlefield actions for the operations research scientist to observe and analyze in order to construct more realistic mathematical models.

b. Provide valid and significant factors and data for input to mathematical models.

c. Provide a means of checking the reality and practicality of the mathematical solutions.

d. Provide evaluations that are not obtainable by other research methods.

e. Be integrated with a computer simulation model to yield information more readily and of greater validity than could be obtained by use of either mathematical models or field experimentation alone.

SOLUTION OF NON-STEADY SOIL MOISTURE TRANSFER

WILLIAM D. OHMSTEDE
U. S. ARMY ELECTRONIC PROVING GROUND
Fort Huachuca, Arizona

The character of non-steady soil moisture transfer has been investigated in conjunction with a basic research program in micrometeorology. An important aspect of micrometeorology is the energy budget at the earth/air interface. Major transformations of the modes of energy exchange are concentrated at the interface. A reckoning of the exchange processes can be made by applying the principle of conservation of energy and thus evaluating the balance that exists between the various modes of energy transport. The more important processes to be considered for thermal energy are net radiation exchange, soil heat conduction, turbulent heat transfer, and evapotranspiration.

The evaporation of water from the soil and transpiration by plants are very important factors to be considered in the energy budget of the earth's surface. Up to 80% or more of the net radiation input at the earth's surface may be used to evaporate water even in an unsaturated field. A quantitative knowledge of evapotranspiration is an essential requirement in micrometeorological models for the surface boundary layer.

With regard to evaporation from fallow soil, the following characteristics can be enumerated:

a. Evaporation from a saturated soil is equivalent to evaporation from an open water surface and is controlled by the meteorological conditions.

b. Evaporation from unsaturated soil is determined almost solely by soil conditions.

c. The transition from atmospheric control to soil control is quite sharp.

Since the earth's land surface is usually unsaturated, it is evident that an independent solution of the energy balance equation at the earth's surface requires knowledge of soil moisture transfer processes.

A major aspect of soil science has been the classification of the soils of the world. However, this classification is based on such factors as color, texture, and genesis, and is of limited application when considering soil moisture transfer. Fortunately, in recent years evaluations of water retention characteristics of soils have become routine. Nevertheless, knowledge of soil moisture transfer characteristics is restricted to only a very few soils. This appalling lack of knowledge has been due in part to inadequate experimental tools and to inadequate physical models of the soil moisture transfer processes. Today, these shortcomings have been overcome, at least to some extent, and the combination of experimental data and numerical results of a plausible physical model will allow the routine determination of the water transfer characteristics of soils.

Soil scientists have long assumed that the transmission of water through soil was analogous mathematically to heat and electrical flow. The analogy with heat flow is superficial but convenient mathematically; for one-dimensional vertical flow, the principle is expressed as follows:

$$q/\rho = K \left[- (1/\rho)\partial \Upsilon/\partial z - g\right] \qquad [1]$$

where q is the upward flux, ρ the density of a fluid, z is the vertical coordinate having increasing magnitude upwards, -g is the acceleration of gravity (directed downward), K is the capillary conductivity, and Υ is the soil moisture tension.

Knowledge of the characteristics of the capillary conductivity (K) is almost entirely restricted to its empirical measurement wherein it is assumed that equation [1] holds. The capillary conductivity varies immensely as a function of the soil moisture tension. For the purpose of this report, it will be assumed that for a given soil the capillary conductivity is a unique function of soil moisture tension. Figure 1 represents the capillary conductivity as a function of soil moisture tension for three soils. It will be further assumed that the soil moisture tension is a unique function of soil moisture content for a given soil. Figure 2 is a representation of the soil moisture content as a function of the soil moisture tension for three soils.

As shown in figures 1 and 2 the capillary conductivity and soil moisture tension vary over wide ranges as a function of soil moisture. For a homogenous soil material equation [1] can be transformed into the following more useful form:

$$q/\rho = -D\partial\Theta/\partial z - Kg \qquad [2]$$

where $D = (K/\rho)\ \partial\Psi/\partial\Theta$ and Θ is the percent moisture of the soil. The parameter D has the dimensions of diffusivity and, except for the gravity term, equation [2] represents an equation equivalent to diffusion of the soil moisture (Θ). For a first approximation solution of soil moisture transfer, it can be assumed that the diffusivity (D) for a given soil is an exponential function of the moisture content. Figure 3 corroborates this assumption and the relationship can be represented by,the following equation:

$$D = D_0 \exp [\beta(\Theta - \Theta_0)] \qquad [3]$$

where D is the diffusivity at moisture content Θ, D_0 and Θ_0 refer to a different diffusivity and moisture content at an arbitrary reference level, and β is a parameter which is characteristic of a given soil material. For this report it is assumed that for a given soil material the diffusivity is a unique function of moisture content as given by equation [3].

Equation [2] is extended to the non-steady case by application of the continuity equation which results in the following:

$$\partial\Theta/\partial t = (\partial/\partial z)\ (D\partial\Theta/\partial z) + g\partial K/\partial z \qquad [4]$$

In general, equation [4] cannot be solved analytically and therefore it is necessary to resort to some form of numerical analysis to find solutions for the various problems of non-steady state soil moisture transfer. This paper presents certain solutions of equation [4] which have applications in determining soil moisture transfer characteristics.

Gardner[(1)] proposed that soil moisture transfer characteristics could be evaluated over a wide range of soil moisture by using the transient outflow data from pressure plate or membrane apparatus. The experimental procedure requires that a nearly instantaneous pressure increase be applied to a soil sample whose tension was initially in equilibrium with the pressure within the pressure plate or membrane apparatus. The outflow of water as a function of time after the pressure step, the total outflow, and the pressure step size are the experimental data required.

In the mathematical treatment, we consider a slab of thickness L and infinite horizontal extent. If we neglect gravity, we can assume that the initial state is in equilibrium and has a constant tension (ψ_1), percent moisture (Θ_1), and diffusivity (D_1). The upper boundary is assumed to be impermeable. At time t=O, the lower boundary comes in contact with a medium of infinite conductivity, infinite capacity, and a tension of ψ_0. Instantaneously, the lower boundary will come to have percent moisture Θ_0 and diffusivity D_0. We now define a new variable v as follows:

$$v = D/D_0 - 1 \qquad [5]$$

Continuing to neglect gravity, application of equations [3] and [5] to equation [4] gives the following:

$$\partial v/\partial t = D_0\,(1 + v)\,\partial^2 v/\partial z^2 \qquad [6]$$

Furthermore, if we change to a dimensionless depth ($x = z/L$) and dimensionless time ($T = D_0\, t/L^2$) equation [6] becomes:

$$\partial v/\partial T = (1 + v)\,\partial^2 v/\partial x^2 \qquad [7]$$

In this notation, the boundary and initial conditions are as follows:

$$\begin{array}{ll} T = 0,\ x > 0 & :\ v = V \\ T > 0,\ x = 0 & :\ v = 0 \\ T \geq 0,\ x = 1 & :\ \partial v/\partial x = 0 \end{array} \qquad [8]$$

Equation [7] is nonlinear and cannot be readily solved analytically. Nevertheless, it is possible to approximate the solution in limited conditions by analytical expressions. For example, if $v << 1$, equation [7] can be approximated by the following linear equation:

$$\partial v/\partial T = \partial^2 v/\partial x^2 \qquad [9]$$

The solution of [9] under conditions [8] is a classical problem of mathematical physics. In our notation the solution is:

$$\frac{v}{V} = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{\sin[(2n+1)\pi x/2] \cdot \exp[-(2n+1)^2\pi^2 T/4]}{(2n+1)} \qquad [10]$$

According to equations [3] and [5], the percent moisture ($\Theta - \Theta_0$) is given by $(1/\beta)\ln(1 + v)$; however, when $v << 1$, $\Theta - \Theta_0 \simeq v/\beta$. Thus, integrating equation [10] from the limits of 0 to 1 gives a measure of the amount of water yet to come out at any time T.

$$Y = \frac{8}{\pi^2} \sum_{n=o}^{\infty} \frac{\exp[-(2n+1)^2\pi^2 T/4]}{(2n+1)^2} \qquad [11]$$

Equation [11] defines the "normalized yet to come" (Y) for equation [9] under conditions [8]. After a very long time, $T > 1$, only the first term of the series is significant and consequently we have the following relationship:

$$\partial \ln Y/\partial T = -\pi^2/4 \qquad [12]$$

Equation [12] is important because it is applicable to the nonlinear solutions of equation [7] at large T as well for v becomes very much less than one and the solution is quasi-linear.

Equations [9 - 12] were suggested by Gardner (1) for determining the soil moisture diffusivity (D_o). This solution is applicable in reality only to the case of infinitesimal step size ($V \ll 1$). In practice, this condition cannot be adequately satisfied, since only an infinitesimal amount of water would be extracted.

Because of these limitations, it is clear that a set of solutions of the nonlinear equation [7] would be very useful for it would allow all data from a soil moisture extraction experiment to be utilized to determine the diffusivity (D). Consequently, a program of numerical analysis was carried out to determine a family of solutions of equation [7] under conditions [8]. A grid of points was established across the slab. The spatial derivatives at each grid point were estimated from finite difference formulae, based upon Lagrangian fourth-order polynominals. Having approximated the spatial derivatives at the grid points by finite difference formulae, equation [7] becomes a set of ordinary first-order equations whose solutions are the values of v at each grid point as a function of dimensionless time (T). The temporal solutions are determined by using a suitable integration routine — in this instance, a fourth-order Runge-Kutta. The normalized yet to come (Y) was evaluated throughout the solution by integration of $\ln(1+v)$ over the slab by the use of suitable integrals of finite difference formulae determined from the grid point data.

Figure 4 summarizes the set of numerical solutions of equation [7] with conditions [8] for $V = 10^4$, 10^3, 10^2, 10, 1, and 0 (linear case). The solutions were run until ln Y exceeded -7.5. According to equation [12], all solutions approach a limiting slope of $\partial \ln Y/\partial T = -\pi^2/4$. Consequently, it was convenient to shift each solution along the yet to

come axis so that all tails in figure 4 match. The multiplying factor by which the solutions are displaced along the ordinate is termed the "shift parameter." Figure 5 is a graph of the relationship of the shift parameter to the step size (V). The shift parameters were determined by matching the solutions, one on another, at the end of each tail.

It will be worthwhile at this point to review how figures 4 and 5 may be used to evaluate soil moisture transfer characteristics. A sample of soil is placed in a container whose ends are open but with walls which prevent lateral flow. The sample is positioned to insure good contact with the underlying ceramic plate or membrane of a soil moisture extractor. The sample is wetted, the extractor is sealed, and pressure is applied to bring about the tension desired for the initial state of the experiment. After sufficient time has been allowed for the soil to come to equilibrium with the pressure within the extractor, the experiment is initiated by rapidly increasing the pressure in the soil moisture extractor by a pressure step (ΔP). Data required are the outflow of water from the soil moisture extractor as a function of time elapsed after the pressure step. Frequent observations are required just after the start. Data must be taken until there is essentially no more outflow. The percent soil moisture extracted (W_∞) is obtained by dividing the volume of water extracted by the volume of the soil sample. The percent accumulative outflow (W_T) is obtained in an analogous manner. The normalized yet to come (Y) is found as follows:

$$Y = (W_\infty - W_T) / W_\infty \qquad [13]$$

The normalized yet to come (Y) is plotted against time following the pressure increase on transparent log-log paper like that of figure 4. This graph is placed over figure 4 and moved, maintaining parallel axes, until the data points fall in logically with the family of curves suggested by those plotted in figure 4. By reading off the ordinate for small time of the curve suggested by the data points, the shift parameter is obtained. The step size (V) is obtained from figure 5 using the shift parameter. The parameter β is found as follows:

$$\beta = \ln(1 + V)/W_\infty \qquad [14]$$

The soil moisture diffusivity (D_o) for the final state is obtained by determining the real time (t_r) which corresponds with the dimensionless time T=1 of the solution and using the following equation

$$D_o = L^2 / t_r \qquad [15]$$

where L is the soil sample thickness. Combining the results of equations [14] and [15] with equation [3] will characterize the soil moisture transfer characteristics over the range of the pressure step taken.

The results to this point have not considered two important complicating factors in transient outflow experiments using membrane extractors. The most important of these is the problem of membrane impedance. If the flow impedance of the membrane is comparable to that of the soil slab, a tension differential across the membrane is required to maintain outflow. Consequently the boundary conditions of [8] are incorrect. The result is that the outflow is less than it would otherwise be and the magnitude of V and D determined by Figures 4 and 5 and equation [15] are underestimated. A linear equation is available to describe this problem but, as before, it is applicable only for infinitesimal step size. Consequently, it was again necessary to obtain numerical solutions to characterize the problem. Twelve solutions were obtained for various prescribed conditions. The details of the problem are rather complex and cannot be presented here; the results are summarized below.

The results presented in figure 4 neglect any effects of gravity on the outflow. The mathematical problem which includes the gravity term is extremely nonlinear and can only be approached by numerical means. Some 18 numerical solutions of the transient outflow problem with gravity were obtained for various conditions. The results are summarized below.

As a consequence of this mathematical investigation of transient soil moisture transfer, the following conclusions have been derived:

a. The linear theory is significantly inaccurate in representing transient flow of water in soils for finite perturbations. For the step-function transient outflow experiments performed for determination of soil moisture transfer characteristics, the numerical solutions presented in this report allow the use of all data obtained from the experiments to be utilized in characterizing the soil moisture diffusivity over the full-range of the step provided there are not significant membrane impedance or gravity effects.

b. When there are significant membrane impedance effects, the character of the nonlinear solutions is very complex. The solutions enumerated in this report are too few in number to fully characterize this problem except when the ratio of membrane impedance to the soil slab impedance at the final state is less than 1/10 and the experimental step size is kept small.

c. Problems with membrane impedance in transient outflow experiments can be minimized by increasing the soil slab thickness, working at higher soil moisture tensions, and keeping the step size small.

d. Under most circumstances gravity effects are negligible or can be readily avoided in transient outflow experiments. This problem can be minimized by decreasing the soil slab thickness, by working at higher soil moisture tensions, and possibly by increasing the step size.

e. All the numerical solutions of transient soil moisture transfer presented in this report are based on the assumption that the soil moisture diffusivity for a given soil is a unique exponential function of the percent soil moisture. Even if this assumption is not valid, the solutions presented in this report should remain reasonably accurate for the small step sizes. More important, however, is the fact that these numerical solutions can be used to verify or determine the limitations of this useful hypothesis.

f. In general, it is concluded that the results of this report are useful and it is hoped that they will promote a more vigorous pursuit of the important problem of determining the soil moisture transfer characteristics of the many soil types of the world.

REFERENCES:

(1)Gardner, W. R. "Calculation of Capillary Conductivity from Pressure Plate Outflow Data," Soil Science Society of America Proceedings, Vol 20, pp 317-320 (1956)

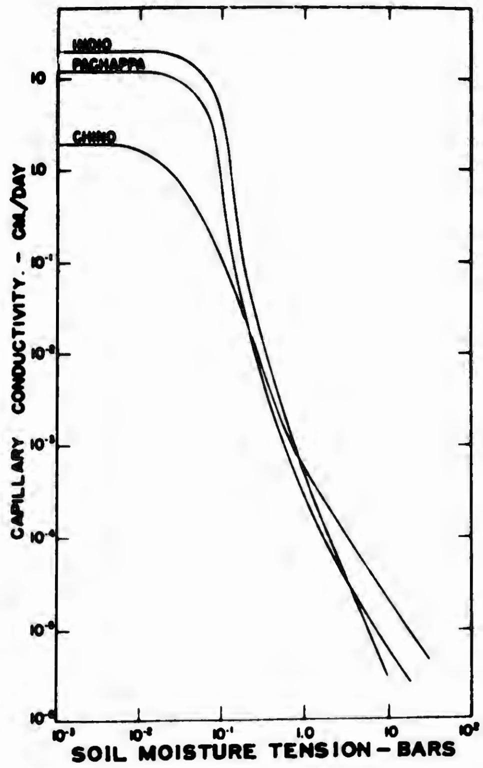


Figure 1. The capillary conductivity as a function of soil moisture tension for three soils.*

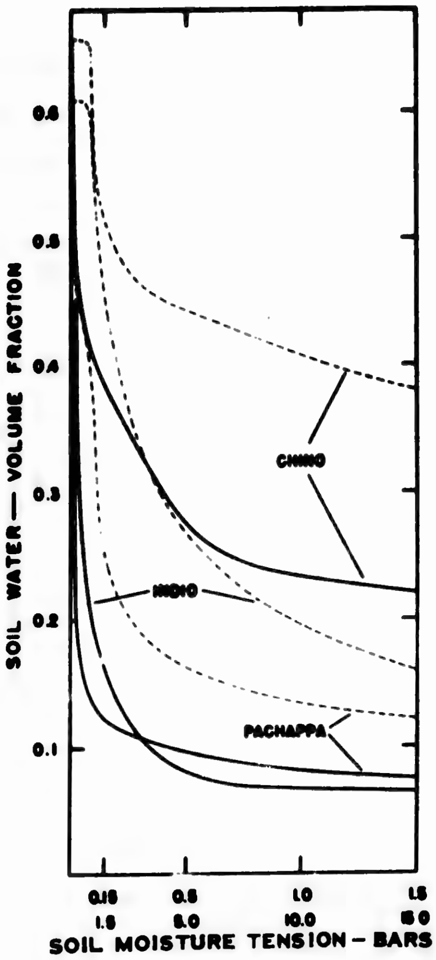


Figure 2. The soil moisture tension as a function of percent soil moisture for three soils.*

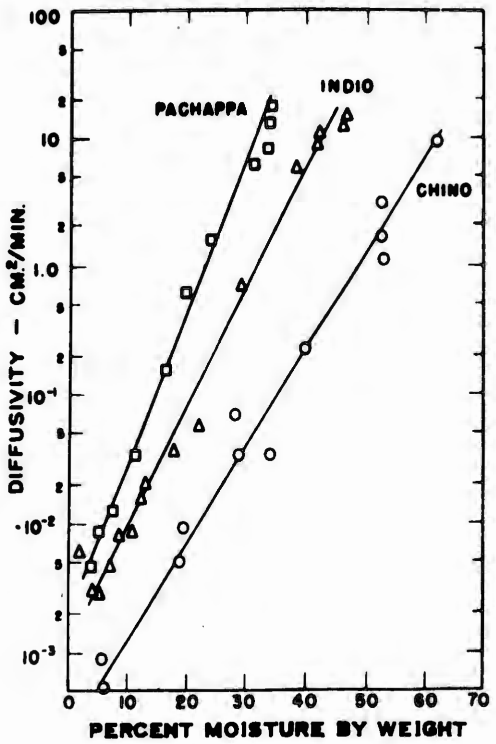


Figure 3. The soil moisture diffusivity as a function of percent soil moisture for three soils.*

*(after Gardner)

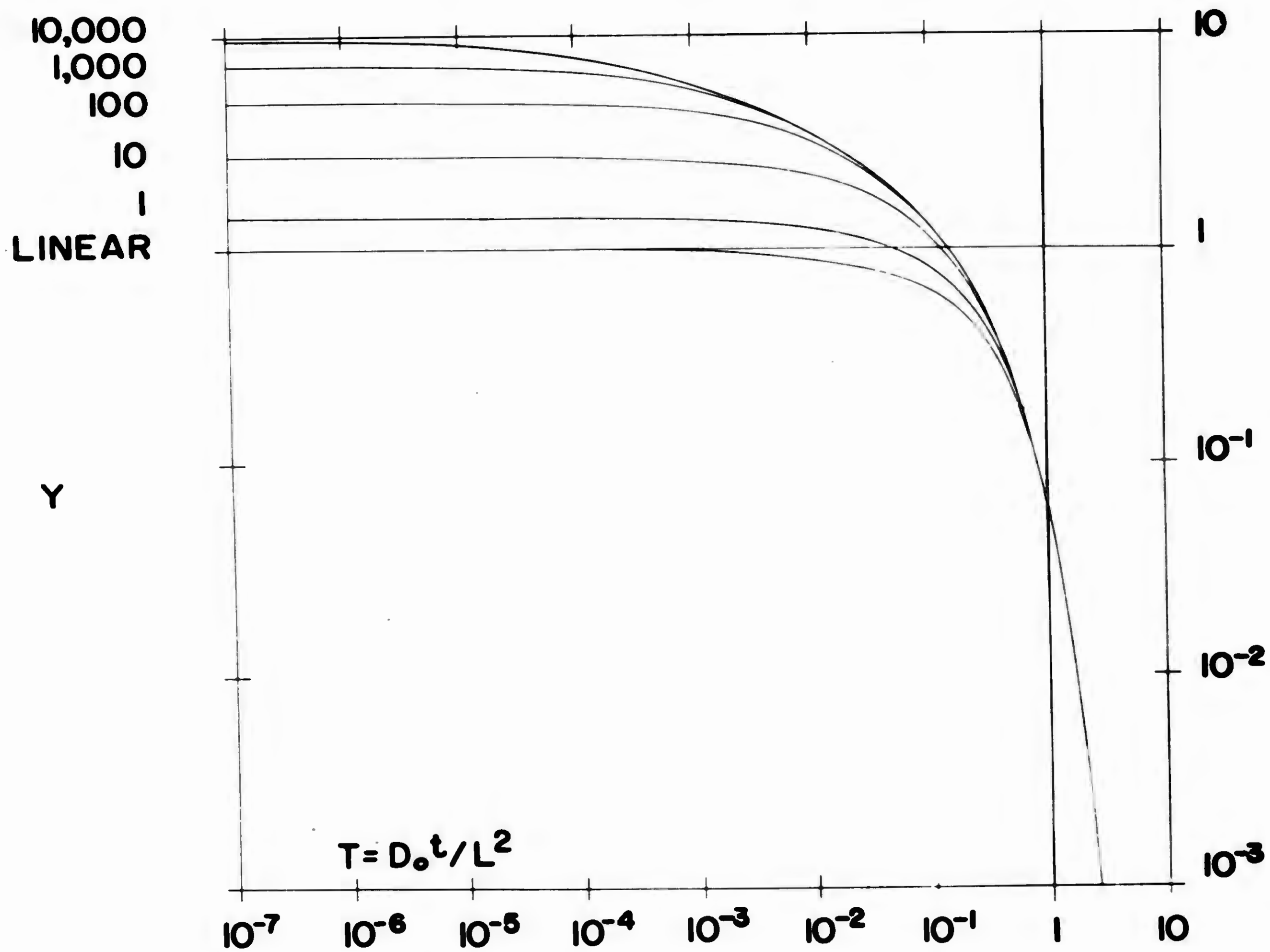


Figure 4. Numerical solutions for various step sizes showing the normalized yet to come (Y) as a function of dimensionless time (T). The curves are shifted along the ordinate in order to match the tails.

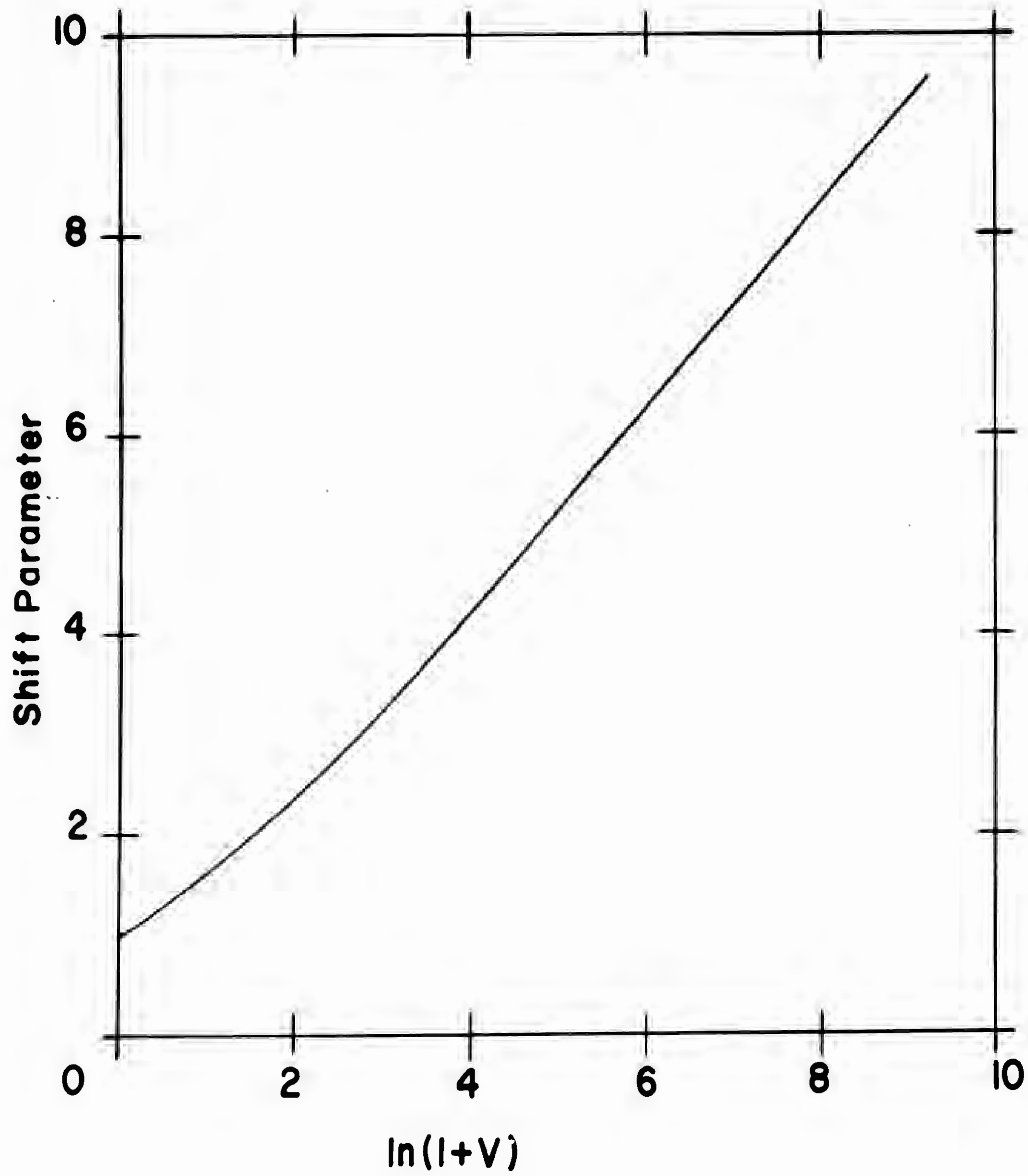


Figure 5. Shift parameter as a function of step size (V).

EFFECTS OF MUSTARD GAS* ON T2 BACTERIOPHAGE AND DNA* SYNTHESIS

B. PAPIRMEISTER
USA CHEMICAL RESEARCH AND DEVELOPMENT LABORATORIES
ARMY CHEMICAL CENTER, MARYLAND

The investigation of mustard gas has covered more than half a century and has engaged the efforts of many investigators in many laboratories. Initially, interest was directed to the vesicant and protoplasmic poisoning actions of this compound. But in the past 15 years, the scope of its interest has been broadened considerably by the discovery that H and other alkylating agents are mutagens[1,2] and anticarcinogens[3,4,5] and have biological actions similar to those produced by ionizing radiations. In spite of repeated observation and extensive review[5,6,7,8,9] of these effects, the mechanism of cellular toxicity of H is still unknown.

While the elucidation of the mechanism of action of H is an enormous task, it must precede the logical development of a therapy for the mustard burn and the reversing of systemic intoxication with the agent, problems of continuing concern to the Chemical Corps. Although there are certain gross similarities between the vesicant action of H on skin and the radiomimetic behavior of H on other cells and organisms, much more must be learned before a valid correlation between the two can be made. With the goal of augmenting the limited knowledge concerning the radiomimetic actions of H, the authors undertook the present investigation of the mechanism of inhibition of T2 bacteriophage by H.

---

* Mustard gas ($Cl-CH_2-CH_2-S-CH_2-CH_2-Cl$) is also known as sulfur mustard and yperite, and it will be denoted in the present report by the symbol H.

DNA--deoxyribonucleic acid.

The present discussion is concerned mainly with matters of general importance for elucidating the mechanism of the inhibition of T2 bacteriophage by H and also includes some observations on the toxicity of the agent for DNA synthesis *in vivo* and *in vitro*.**

Summary of Effects of H on T2 Bacteriophage

The effects of H on some properties of the T2 bacteriophage system are graphically summarized in figure I. To provide a more meaningful basis for comparing the sensitivities of H of the various properties, the information in figure I is divided into two categories: phage particles that were H-treated before infection (dashed lines) and after infection (solid lines). This distinction is necessary because extracellular H-treatment produces secondary damages that prevent some of the particles from penetrating susceptible hosts and limits the progress of the intracellular events of the infective cycle one may wish to investigate. Although some of these complications are overcome when particles are exposed to H after adsorption (when many of the H-sensitive steps associated with penetration are already completed), the interpretation of results obtained with phage-bacterium complexes is more difficult because of the concomitant exposure of the host cell. Recognizing that exact quantitation is not possible in every instance, the authors feel that certain meaningful conclusions nevertheless can be drawn.

By examining the relative slopes of the inactivation curves in figure I, one will note the following features with respect to the action of H on the T2 bacteriophage system:

1. When singly infecting bacteria, the abilities of T2 bacteriophage to reproduce (curves 1 and 3) and to synthesize viral DNA (curve 2) are equally inhibited by H and are the most H-sensitive properties found.

2. Most of the properties of the virus particle directly or indirectly attributable to functions of viral protein are inhibited by H to similar extents (curves 4, 5, 6, 7, and 8). An exception is the ability of the particle to adsorb on susceptible hosts (curve 15). However, these properties are much less sensitive than the ability of the virus to reproduce (curve 1).

** For experimental details the reader is referred to CRDL Special Publication 2-45, October 1961, "On the Mechanism of Inhibition of T2 Bacteriophage by Mustard Gas."[10] Reprints are available from the author on request.

3. The ability of singly infected cells to reproduce T2 phage is much less sensitive to H when the agent is administered 9 minutes after infection (curve 9) than when applied at an earlier stage (curve 3).

4. Limited mainly by the action of H on the cell's capacity to support phage development (curve 12), the abilities of multiply infected cells to reproduce T2 bacteriophage (curve 10) and to synthesize viral DNA (curve 11) are fairly resistant to H.

5. The sensitivity to H of a genetic marker, i.e., the ability of an H-inactivated parental particle to contribute the marker to progeny in a mixed infection with viable carrier phage, is only about 8% of the sensitivity of the particle to reproduce (compare curves 13 and 3).

6. The ability of an H-inactivated particle to initiate the biosynthesis of an enzyme characteristic of the infected cell is only about 5% as sensitive to H as is the ability of the particle to reproduce (compare curves 14 and 3).

The possible meaning of these and some other findings and their bearing on the mechanism of action of H will now be discussed.

The Site of Action of H in T2 Bacteriophage

The current findings support the idea that the most H-sensitive target in T2 bacteriophage is viral DNA and that protein functions are also affected. The evidence is the following:

1. Particles are rapidly inactivated by low doses of H either when free or when preadsorbed on host cells (figure I, curves 1 and 3), indicating a relatively small effect on protein functions. This conclusion was reinforced by determining that individual properties associated with the viral protein such as killing ability, injection of DNA, adsorption, etc., are much less sensitive to H than the ability of the virus to reproduce (figure I, curves 4, 5, 6, 7, 8, and 15).

2. At an H dose corresponding to 1 phage-lethal-hit, a minimum of 45 molecules of H react with viral DNA and approximately 30 molecules alkylate viral protein (table I).

3. H-treated DNA is unstabilized to heating at 74°C, a temperature that does not affect normal DNA.[10]

4. The capacity of the bacterium to support phage reproduction and to synthesize viral DNA is not especially sensitive to H (figure I, curve 12), showing that the enzymic machinery of the host is not seriously impaired by the agent.

5. Vegetative phage becomes abruptly resistant to H after 6 to 9 minutes of metabolism (figure I, curve 9), resembling findings with other agents that are thought to act on viral DNA (ultraviolet or X irradiations, decay of incorporated P^{32}, etc.).

6. Phage particles that contain H-treated DNA but functional protein are noninfective. Such particles are in many ways similar to particles that have been exposed to H while intact.[10]

7. H produces damages that are localized in the genetic material of T2 and T4 bacteriophages. An individual marker is about 8% as sensitive as is the ability of the particle to reproduce (figure I, curve 13).

8. H-inactivated T2 particles can be made to multiplicity reactivate with high efficiency, showing that a genetic reconstitution process can be invoked to reconstitute a viable genome from undamaged portions surviving the H-treatment.[10]

9. H significantly modifies the priming activity of DNA in the *in vitro* DNA synthesizing reaction catalyzed by *E. coli* polymerase (table II). Since DNA synthesis is also inhibited by H *in vivo* (figure I, curve 2), the finding provides independent evidence that H acts on a "template" function of DNA.

The chemical reaction(s) of H with DNA that may actually be responsible for toxicity of the agent is(are) not clear at the present time. Although alkylations of guanine, adenine primary phosphate groups, etc., have been reported,[11, 12, 13, 14, 15] Rutman, *et al*.,[16] suggest that the extent of alkylation probably does not bear any quantitative relation to the biological effectiveness of mustards and that the type of reaction appears to be more important than the extent. This view is supported by the author's finding (table I) that, with an H dose sufficient to produce one phage-lethal-hit, there occur 40 or more combinations per phage equivalent of DNA. It is probable that only one of these combinations is responsible for toxicity and that, consequently, most of the other reactions are nonlethal, although some of them may "mutate" the DNA molecules.

Polyfunctional vs. Monofunctional Agents

Polyfunctional mustards of both the sulfur and nitrogen types are several orders of magnitude more toxic to T2 bacteriophage than the monofunctional mustards, supporting a cross-linking mechanism.[10] Aside from considerations of dose, the two agent types appear to differ also in the mechanism by which they inhibit phage reproduction. Polyfunctional agents cause only very localized lesions in the phage genome and leave the greater portion of the genome intact, available for rescue by genetic recombination processes such as multiplicity and cross-reactivations.[10] On the other hand, some

single-armed agents inactivate phage particles largely by inhibiting protein function that are essential for infection. A single-armed mustard that acts in a different way is β-chloroethyldiethylamine (semi-HN). While still retaining its preferential action on viral DNA and in this respect resembling the action of a polyfunctional agent, it produces lethal damages to phages that are, however, not multiplicity reactivatable (unpublished observation). This finding suggests that the lethal damages produced by semi-HN, in contrast to those produced by the polyfunctional agents, may not be highly localized within the genetic structure and that viral DNA might be inactivated only after extensive reactions involving the greater portion of the genome.

Although cross-alkylation of DNA appears to be an important first step in inhibiting the reproduction of T2 bacteriophage, a different mechanism is probably responsible for other biological actions of the mustards. Single-armed agents are very efficient as mutation producers,[1, 2, 17] and also retain vesicant properties.[18] There is limited evidence that some of these effects may also originate from an initial attack of the agent on DNA. Thus one could suppose that a single alkylating attack (by a single-armed agent or by a half-reacted double-armed agent) may modify a base such as guanine, altering its base-pairing characteristics, and in this way producing a mutation.[15, 17] A single alkylation of a base in DNA probably does not affect the structural integrity of the macromolecule but merely modifies the genetic information, whereas a single cross-link would produce a profound change in DNA, putting the molecule out of action and finally killing the organism.[11]

Unfortunately, little can be said at present regarding the underlying basis for the vesicant action of the mustards and, indeed, a great deal more information than is now available is required to relate a particular biological effect of H to a specific reaction of H with cellular components.

Interaction of H-treated DNA with Biologically Important Proteins

An indication that H affects the structural integrity of the DNA molecule is the observation that H-damaged DNA loses its ability to associate with various types of proteins. The evidence that protein attachment to DNA is impaired by H was derived from 3 different experiments, and a biologically important process was inhibited in every case. The results of these experiments were:

1. When viral protein synthesis was resumed following treatment with a moderate dose of H of phage precursor DNA (formed in the absence of protein synthesis), intact particles were not formed.[10] In similar experiments employing ultraviolet irradiation instead of H, Tomizawa[19] observed efficient production of finished phages, although most of them were noninfective. These results demonstrate that,

although UV-irradiated DNA can, H-treated DNA cannot, interact productively with viral protein.

2. When singly infected cells are exposed to H 9 minutes after infection, the ability of the virus to reproduce is more resistant to H than when the agent is administered at an earlier time (compare figure I, curves 9 and 3). Uchida and Stent[20,21] obtained a similar result with phages that were killed by decay of incorporated P^{32}, and they showed that the acquisition of resistance at the ninth minute might be due to the association with viral DNA of a newly synthesized protein. It is probable that a similar mechanism is responsible for protecting viral DNA from being lethally damaged by H when the agent is applied at the ninth minute. However, no protection is afforded at the earlier times when the protective protein has not yet been synthesized. The question of whether H-inactivated particles are capable of synthesizing the protective protein has not been answered, but it should be recalled that such particles do initiate the biosynthesis of several enzymes (figure I, curve 14). Even if the protective protein were similarly synthesized, it might fail to associate with H-damaged DNA and thus fail to protect.

3. The ability of DNA to prime the _in vitro_ synthesis of DNA that is catalyzed by _E. coli_ polymerase is significantly modified by H (table II, also see reference 10). When H-treated DNA was used as primer, DNA synthesis was inhibited for a time but was then renewed at a rate similar to that observed with untreated DNA. The findings suggest that H-treated DNA has a lower affinity for polymerase than does normal DNA. If it should develop that the association of DNA and polymerase are also impaired by H _in vivo_, the mechanism of the temporary inhibition of DNA synthesis would be explained. Although the DNA synthesis finally recovers, cells would be killed because of "unbalanced growth."[22]

Some Genetic Consequences

A significant proportion of the H-induced lethal damages to T2 bacteriophage appear to be located in the genetic structure of the particle. When particles are exposed to the agent before adsorption, only a maximum of 30% of the lethal damages are sustained by viral protein, some of which may prevent the particle from penetrating the host bacterium (figure I, compare curves 1 and 8). When particles are treated after adsorption, only a very small fraction of the lethal damages result in inhibition of the cell's capacity to support phage development (figure I, compare curves 3 and 12). These results suggest that the majority of the lethal damages to T2 bacteriophage probably occur in the viral genome, and the author was able to confirm this conclusion by more direct evidence. Results from marker rescue experiments show that H inactivates individual genetic markers and that the sensitivity of a given marker is about 8% of the sensi-

tivity of the whole genome (figure I, compare curves 3 and 13). Because several markers at different sites within the genome are about equally affected, H seems to produce genetic damages randomly along the chromosomal structure. The highly localized nature of the genetic lesions produced by H means that inactive particles contain considerable stretches of undamaged genetic material, a notion that is compatible with the observed high efficiencies of multiplicity and cross reactivations.[10] It should be stressed that such reactivations probably are not due to chemical reversal of H damages but rather are due to genetic mechanisms that are capable of reconstituting a viable genome from undamaged genetic sections derived from several parents. A similar mechanism can also reactivate particles that have been killed by ultraviolet or X irradiations, agents that produce their lethal damage by different reactions.[23]

The ability of H to produce mutations was not investigated in the present work. However, in a recent report Bautz and Freese[17] indicated that the ethylation by alkylating agents (presumably of the 7-N position of guanine) of DNA from T4 bacteriophage increased the mutation rate. These investigators suggested that the eventual elimination from DNA of a base by secondary hydrolysis of the N-glycosidic bond might have been responsible for the mutation. They further hinted that reactions with primary phosphoryl groups of the DNA backbone proably do not have mutagenic consequences for they would fracture the molecule by breaking the sugar phosphate bond that links two adjacent nucleotides. Although such a mechanism might be involved when lethal DNA damages are produced by the decay of incorporated P^{32}, it has not been demonstrated to be mutagenic. It is interesting that Bautz and Freese found single-armed alkylating agents to be quite potent as mutagens while the author showed that similar monofunctional mustards are not nearly as toxic to these same particles as polyfunctional compounds. It therefore appears that alkylating agents can modify the DNA molecule in at least two distinct ways: one type of reaction, e.g., a single alkylating attack on guanine, leads to a point mutation; and the other type of reaction, e.g., a crossalkylating attack by a polyfunctional mustard, culminates in a lethal event and entirely inactivates the affected gene.

Some Metabolic Consequences

1. The effect of H on the host cell and its capacity to support phage development.

The most H-sensitive properties of *E. coli* B bacteria are the abilities of cells to reproduce and to synthesize DNA. However, H-treated bacteria are still able to increase in size,[24] synthesize RNA and protein,[25] carry out respiratory metabolism, and produce adaptive enzymes.[26] Coupled with the observed inhibition of DNA synthesis (which may be only temporary[25]), the otherwise normal metabolism apparently leads to an imbalance, which may ultimately

kill the cell by "unbalanced growth."[22] The death-by-unbalanced-growth theory is compatible with the author's recent finding (unpublished) that significantly larger bacterial survivals can be obtained by permitting H-treated E. coli B to metabolize at a slower rate. It is suggested that cells have a limited ability to repair or circumvent some (although not all) of the lethal H damages.

Although sterilized by H and unable to synthesize bacterial DNA, E. coli bacteria remain capable hosts for T2 bacteriophage, supporting viral reproduction and synthesizing almost normal amounts of viral DNA.[10] The findings confirm similar observations by Herriott[24] and show that the enzymic machinery of the host cell is not damaged sufficiently by H to prevent it from making its necessary contributions to sustaining the viral infection. Exactly how T2 bacteriophage is able to overcome the block in the DNA synthetic mechanism is not known, but it should be recalled that the virus supplies a new set of functional genes, which establish new synthetic pathways especially geared to the synthesis of phage-specific DNA.[27]

2. The inhibition by H of DNA synthesis *in vivo* and *in vitro*, and its bearing on the toxic mechanism.

The inhibition of DNA synthesis by H may be the metabolic block primarily responsible for the cytotoxicity of the agent. In bacterial cells, even a temporary interruption of the process could lead to the unbalanced growth situation, culminating in the eventual death of the organism.[22] In the T2 bacteriophage system, the ability of the virus to reproduce itself when singly infecting a susceptible host and the capacity of the virus to initiate the synthesis of viral DNA are equally sensitive to H (figure I, compare curves 1 and 2), possibly indicating a causal relationship between these two properties. It is to be noted, however, that the infectivity of the particle and viral DNA synthesis become more resistant to H when the agent is administered at a later stage of infection (figure I, curve 9) or when cells are infected with more than one particle (figure I, curve 11). Although the mechanism presently is not understood, resistance is in some way associated with maintaining the integrity of the parental genome, which may become associated at about the ninth minute with a protective protein or which,[20,21] once damaged by H, can be reconstituted by multiplicity reactivation. The *in vivo* evidence therefore leads to the conclusion that H inhibits DNA synthesis by damaging the "DNA template", and an intact template is need for DNA replication.

H also affects DNA synthesis *in vitro*, a reaction that is catalyzed by E. coli polymerase and which requires a DNA primer.[10] The demonstration (table II) that the priming ability of DNA is the most H-sensitive component of the reaction confirms the notion that the agent interferes with a template function of DNA. The lag before the *in vitro* DNA synthesis finally renews is strikingly

similar to that observed in the proposed reconstitution of an active genome by multiplicity reactivation. The results suggest that the effects of H on the in vitro synthesis of DNA are very similar to the effects observed in vivo and therefore reinforce the belief that a molecular basis for the cellular toxicity of H is the ability of the agent to damage DNA.

Another possibility for explaining the inhibition of DNA synthesis by H would be that the agent, by damaging gene function, prevents the formation of enzymes critical to DNA synthesis. (It should be recalled that the catalytic activities of such enzymes are not especially sensitive to H.[10]) The author demonstrated[10], however, that the ability of T2 particles to initiate the biosynthesis of one such critical enzyme (phage-specific deoxyguanylate kinase) is only 5% as sensitive to H as is the infectivity (figure I, compare curves 3 and 14). This implies that most inactive particles probably are not prevented from carrying out the gene functions associated with enzyme formation. In spite of this discovery and even if H-inactivated particles were similarly capable of forming other enzymes, one could not rule out entirely the possibility that DNA synthesis might be inhibited by a damage to the enzyme-forming apparatus. Thus, one could suppose that inactive particles harbor damages at different genetic sites, remain capable of synthesizing most enzymes, but fail to produce the complete set of enzymes need for DNA synthesis. Such a mechanism might be invoked to explain the phenomenon of multiplicity reactivation, which restores DNA synthesis. In mutliply infected cells, in contrast to singly infected cells, several inactive particles may pool their enzyme-forming potentials to establish a complete enzyme set, and the inhibited viral DNA synthesis recovers (figure I, curve 11).

Current studies are designed to obtain more information on the effect of H on enzyme biosynthesis. Of special concern is the ability of H-inactivated T2 particles to initiate the synthesis of phage-specific polymerase,[27] an enzyme peculiar to the infected cell and primarily responsible for catalyzing the viral DNA synthesis.

3. The sensitivity to H of phage-bacterium complexes at different stages of the latent period.

Four major hypotheses (which are not mutually exclusive) may be offered to explain the development of resistance to H at the ninth minute after single infection: (1) the parental DNA, in matings with progeny DNA, may reconstitute a viable genome by multiplicity reactivation; (2) the parental DNA may associate with a newly synthesized protein, which might protect the DNA from being lethally damaged by an otherwise lethal agent; (3) the genetic information originally carried by the sensitive parental DNA may be transferred to nonsensitive structures, thus making the parental DNA dispensable for sustaining the infection; and (4) all H-sensitive

functions required for DNA replication may be completed by the ninth minute, so the agent is no longer lethal when administered at this time, i.e., all enzymes necessary for viral DNA synthesis are already present at this stage of the latent period.

Although there is some experimental evidence to support each of the 4 proposals, one cannot presently attribute the 9-minute resistance to any one, or any combination, of them.

Multiplicity reactivation is an efficient process when multiply infected bacteria are exposed to H at a very early stage of development, but has not yet been shown to occur when the agent is administered at the ninth minute. Also, whereas a normal multiplicity reactivation process involves matings between several parental particles, in the present instance matings would have to be between a parental and one or more progeny genomes or between several of the progeny. It should be stressed that these are unsolved problems and do not eliminate multiplicity reactivation as a possible agency responsible for making monocomplexes resistant to H at the ninth minute.

The formation by infected cells of a protein capable of protecting parental T2 particles from being killed by the decay of incorporated P^{32} has been inferred from studies by Uchida and Stent.[20,21] These investigators also demonstrated that this protein protects specifically only those parental particles responsible for its formation. It does not protect secondarily infecting particles they would have to synthesize their own protein to become resistant to P^{32} decay. These results may mean that such protein becomes physically associated with the DNA of the parent particle. It may exert its protective action by bracing the backbone of parental DNA and thus preventing a complete scission of the macromolecule by P^{32} decay or, possibly, by other agents. Another possibility is that the protein protects DNA by separating the complementary strands (preparatory to DNA synthesis, which begins at approximately this time), thus preventing simultaneous damages from occurring at complementary sites of the duplex (i.e., a damage to only one of the two complementary sites may not be lethal since the genetic information is still available on one strand). To test some of these notions, current investigations are under way in this laboratory to determine if polymerase can protect the ability of DNA to prime the *in vitro* synthesis of DNA from being damaged by H.

Little can be said regarding the proposed transfer of genetic information from the H-sensitive parental DNA to H-resistant molecules. The most logical candidate substance would be RNA which, because of its structural differences from DNA, is more resistant to H.[28] Although the synthesis of "phage-specific RNA," which mimics the base composition of parental DNA, has been reported[29] and is apparently required for phage reproduction, such RNA has not been

shown to be a template for the synthesis of progeny DNA. Recent *in vitro* findings also point to a definite role of DNA in the synthesis of RNA; but, again, a requirement for an RNA primer for DNA synthesis has not been found. These results suggest that phage-specific RNA represents an aspect of DNA function rather than DNA replication, and it may be an important intermediate in the translation of genetic information from parental DNA into the amino acid sequence of a protein or enzyme. Proteins that are generally more resistant to H and similar agents than are the nucleic acids are not well suited for carrying the genetic code and will not be considered here as an intermediate template in the synthesis of progeny DNA. However, the denial of a genetic role for proteins should in no way minimize the importance of these molecules as catalysts for DNA or, perhaps, even as having a function in the determination of the size of the DNA polymer.

An intriguing possibility is that the resistance at the ninth minute may be due to the fact that all the enzymes necessary for viral DNA synthesis are present by this time. When further protein synthesis is inhibited by chloramphenicol added at the ninth minute, subsequent DNA synthesis is not appreciably affected although the antibiotic does prevent the initiation of DNA synthesis when given at an earlier time.[30] The similarities in the time spectrum relative to the action of chloramphenicol (an agent that inhibits protein synthesis) and of the radiations or H (agents that affect DNA directly) lead one to suspect that the resistance of monocomplexes to H is also related, in an as yet obscure manner, to an inhibition of the biosynthesis or function of one or more of the important proteins or enzymes required for viral DNA synthesis. In this connection it was found, however, that most H-inactivated T2 particles retain their ability to initiate the biosynthesis of several of such enzymes.[10] Current studies are designed to investigate further the biosynthesis of proteins and their importance for the development of the H-resistance at the ninth minute.

Regardless of mechanism, the discovery that not all stages of phage development are equally sensitive to H emphasizes the importance of studying the action of a nucleotoxic substance in a synchronized system. Perhaps some of the conflicting results in the literature relative to the effects of H in other systems are due to a failure to work with synchronized cell populations.

4. The ability of H-inactivated T2 particles to initiate the biosynthesis of enzymes and the importance of this phenomenon for relating gene function to gene replication.

Following infection of *E. coli* B with T2 bacteriophage, several new enzymes appear that are not present in uninfected cells.[26] It appears likely that these enzymes represent products of the

phenotypic function of viral DNA. Even in single infection (i.e., in the absence of multiplicity reactivation) H-inactivated phage are capable of initiating the biosynthesis of several of these enzymes. For example, when enzyme activity was assayed 15 minutes following infection, the ability of the particle to synthesize phage-specific deoxyguanylate kinase[31] was never found to be greater than 5% as sensitive to H as were the abilities to reproduce and to synthesize DNA (figure I, compare curves 2, 3, and 14). Whether or not these enzyme assays quantitatively reflect the biosynthetic potentials of H-inactivated virus, it is apparent (assuming that these enzymes are products of gene function) that gene function precedes gene replication and may even be manifested in its absence.

The separation of the 2 most important attributes of genetic material, gene function and gene replication, may explain the nucleotoxicity of H. To wit, most H-treated cells, although sterile, continue to grow in size, synthesize protein and RNA, and carry out almost normal respiratory activities. At least some of these metabolic processes are directly or indirectly controlled by the genetic material. While a single H damage to DNA is sufficient to inhibit the reproduction of all the genetic material, thus preventing cell multiplication, the remaining undamaged portions of the genome continue to carry out their phenotypic functions, providing the cell with almost all the factors needed for growth.

The finding that gene function precedes gene replication (all enzymes needed for viral DNA synthesis are present by the ninth minute) may mean that double-stranded DNA is responsible for phenotypic action. Contrariwise, soon after parental DNA becomes single-stranded (at about the ninth minute), DNA synthesis begins and enzyme biosynthesis ceases. The possibility that enzyme biosynthesis is therefore controlled by the state of the parental DNA is supported by the discovery that particles killed by UV irradiation cannot synthesize DNA but continue to produce enzymes for much longer periods and in much larger amounts than do viable phages.[32] The author is presently studying these phenomena using H-treated particles.

It is of interest that the ability of a single genetic marker to reproduce itself in matings with viable carrier phage (marker rescue experiments) was only slightly more sensitive to H than was the ability of the particle to initiate the synthesis of an enzyme (figure I, compare curves 13 and 14). This finding suggests that a single lethal event may be responsible for inhibiting gene replication and gene function, and that a particle harboring such a lethal damage is rendered sterile. However, the ability of the particle to reproduce itself can be restored when damaged genetic sections are replaced by functional components through matings among several particles (multiplicity or cross-reactivations). Thus, for phage viability, one requires a complete set of gene functions (enzymes, phage proteins, etc.) as well as at least one intact genome to serve as a primer for its own duplication.

Reversal of the H-Damages

There are a number of indications that E. coli B bacteria, T2 bacteriophage, and DNA have a limited capacity for repairing or circumventing H-induced damages to the replicating mechanisms. The survival of H-treated E. coli B partially depends on the medium on which organisms grow following exposure. Unpublished results demonstrate that highest bacterial survivals are obtained on a purely synthetic medium (inorganic salts and glucose) and lowest survivals on a nutritionally rich medium (yeast extract). In addition, chloramphenicol added to either medium has a small beneficial effect on survival, but a mixture of ribonucleotides (although not of deoxyribonucleotides) added to the synthetic medium lowered the survival of H-treated cells. Generally speaking, the findings suggest that when the growth rate of H-treated bacteria is slowed, survival is favored by delaying the onset of "unbalanced growth"[22] while permitting some repair of the H-damages to proceed.

T2 bacteriophages can efficiently repair the lethal H-damages by multiplicity reactivation when several inactive particles pool their undamaged genetic segments to reconstitute a new viable genome. Furthermore, T2 bacteriophage (even in single infection) becomes significantly more resistant to H after some 6-9 minutes of intracellular development, indicating that particles have passed some H-sensitive stage or are now capable of repairing or circumventing the H-damages. All chemotherapeutic attempts on H-inactivated T2 particles were unsuccessful, although some of the compounds used (notably aminoethylthiouronium bromide [AET] and mercaptoethyl amine [MEA]) had a prophylactic, dose reducing action when administered prior to or simultaneously with, H.

H-treated DNA spontaneously reacquires the ability to prime the in vitro synthesis of DNA which is catalyzed by E. coli B polymerase.[10] A similar lag before DNA synthesis renews spontaneously has also been reported in vivo with H-treated E. coli.[25] It is proposed that, with the aid of polymerase, new and functional primers are reconstituted from the polydeoxynucleotide sections that were produced by the action of H on DNA. The finding that repair of the H-induced inhibition of the ability to synthesize DNA occurs spontaneously is encouraging from a therapeutic standpoint, although it must be remembered that such newly synthesized DNA has not been analyzed for genetic competence.

Probably the potentially most fruitful approach to eventual therapy of H-induced lesions to cells and viruses is the elucidation of the biochemical mechanisms responsible for the H-resistance of certain mutant strains of E. coli B. Since mutants such as E. coli B/r are similarly resistant to all agents that are presumed to exert their toxic effects through an action on DNA, the molecular basis for resistance is presently quite obscure. On the other hand, retaining

the basic assumption that DNA is the primary target, it is feasible that damaged DNA might be secondarily degraded by nucleases, which are present in sensitive,but may be absent in resistant strains, thus permitting repair to take place in the resistant organisms. A similar mechanism was invoked to explain the author's discovery that the priming ability of heated H-treated DNA does not recover when impure *E. coli* polymerase is used to catalyze the *in vitro* synthesis of DNA.[10] Presumably, the contaminating phosphodiesterase is able to degrade H-damaged DNA. However, when the contaminating phosphodiesterase activity is reduced by further purification of polymerase, the priming ability of H-treated DNA recovers after a lag. Recent findings indicate that secondary breakdown of DNA also occurs with H-treated *E. coli* B and *E. coli* $15T^-A^-U^-$ bacteria (sensitive strains) but is somewhat less with H-treated *E. coli* B/r organism (resistant strains). With the hope of clarifying the biochemical basis for resistance by some mutants of *E. coli*, the author is presently investigating nuclease levels in these organisms.

BIBLIOGRAPHY

1. Stevens, C. M., and Mylroie, A. Biochim. et Biophys. Acta 8, 325 (1952).
2. Auerbach, C. Ann. N. Y. Acad. Sci. 68, 731 (1958).
3. Farber, S., *et al*. Greenstein, J. P., and Haddow, A. (eds.) Advances in Cancer Res. 4, 1 (1956).
4. Stock, C. C., Greenstein, J. P., and Haddow, A. (eds.) Advances in Cancer Res. 2, 425 (1954).
5. St. Whitelock, O. V. Ann. N. Y. Acad. Sci. 68 868 (1958).
6. Philips, F. S. J. Pharmacol. Exptl. Therap. 99, Part II, 281 (1950).
7. Butler, J. A. V. Radiation Res. Suppl. 1, 403 (1959).
8. Timmis, G. M. Biochem. Pharmacol. 4, 49 (1960).
9. Butler, J. A. V., Gilbert, L. A., and Smith, K. A. Nature 165, 714 (1950).
10. Papirmeister, B. CRDL Special Publication 2-45, October (1961).
11. Stacy, K. A., *et al*. Ann. N. Y. Acad. Sci. 68, 682, (1958).
12. Reiner, B., and Zamenhof, S. J. Biol. Chem. 228, 475 (1957).
13. Walker, I. G., and Watson, W. J. Can. J. Biochem. and Physiol. 39, 377 (1961).
14. Brookes, P., and Lawley, P. D. Biochem. J. 80, 496 (1961).
15. Lawley, P. D., and Brookes, P. Nature 192, 1081, December (1961).
16. Rutman, R. J., Steele, W. J., and Price, C. C. Biochem. and Biophysical Res. Comm. 4, 278 (1961).
17. Bautz, E., and Freese, E. Proc. Natl. Acad. Sci. U. S. 46, 1585 (1960).
18. Du Vigneaud, V., *et al*. J. Am. Chem. Soc. 70, 1620 (1948).
19. Tomizawa, J. Virology 6, 55 (1958).
20. Uchida, H., and Stent, G. S. J. Mol. Biol. 2, 251 (1960).
21. Uchida, H., and Stent, G. S. J. Mol. Biol. 2, 262 (1960).

22. Cohen, S. S., and Barner, H. D. Proc. Natl. Acad. Sci. U. S. 40, 885 (1954).
23. Stent, G. S. Adv. in Virus Res. 5, 95 (1958).
24. Herriott, R. M. J. Gen. Physiol. 34, 761 (1951).
25. Harold, F. M., and Ziporin, Z. Z. Biochim. et Biophys. Acta 26, 492 (1958).
26. Marmur, J., and Doty, P. Nature 183, 1427 (1959).
27. Kornberg, A., et al. Proc. Natl. Acad. U. S. 45, 772 (1959).
28. Herriott, R. M. J. Gen. Physiol. 32, 221 (1948).
29. Volkin, E., and Astrachan, L. Virology 2, 149 (1956).
30. Hershey, A. D., and Melechen, N. E. Virology 3, 207 (1957).
31. Bessman, M. J., and Van Bibber, M. J. Biochem. and Biophys. Res. Comm. 1, 101 (1959).
32. Dirksen, M. L., Wiberg, J. S., Koerner, J. F., and Buchanan, J. M. Proc. Natl. Acad. Sci. U. S. 46, 1425 (1960).

LEGEND FOR FIGURE I

1 - Reproduction of T2 following single infection.

2 - Viral DNA synthesis by singly infected cells.

3 - Reproduction of T2 by singly infected cells.

4 - Lysis of bacterial hosts by phage ghosts.

5 - Ability of phage to kill host cells.

6 - Ability of phage to exclude other phages.

7 - Ability of phage to inhibit bacterial RNA synthesis.

8 - Ability of phage to inject its DNA.

9 - Sensitivity of singly infected cells 9 minutes after infection.

10 - Reproduction of T2 by multiply infected cells.

11 - Viral DNA synthesis by multiply infected cells.

12 - Capacity of bacteria to support phage reproduction.

13 - Sensitivity of a host-range genetic marker.*

14 - Ability to initiate the biosynthesis of an enzyme peculiar to the infected cell.*

15 - Adsorption on host bacteria.

---

* These data are corrected for loss of bacterial capacity. The multiplicity of infection was 0.1 T2 particle per cell.

Figure I

SUMMARY OF THE EFFECTS OF H ON PROPERTIES OF THE T2 BACTERIOPHAGE SYSTEM

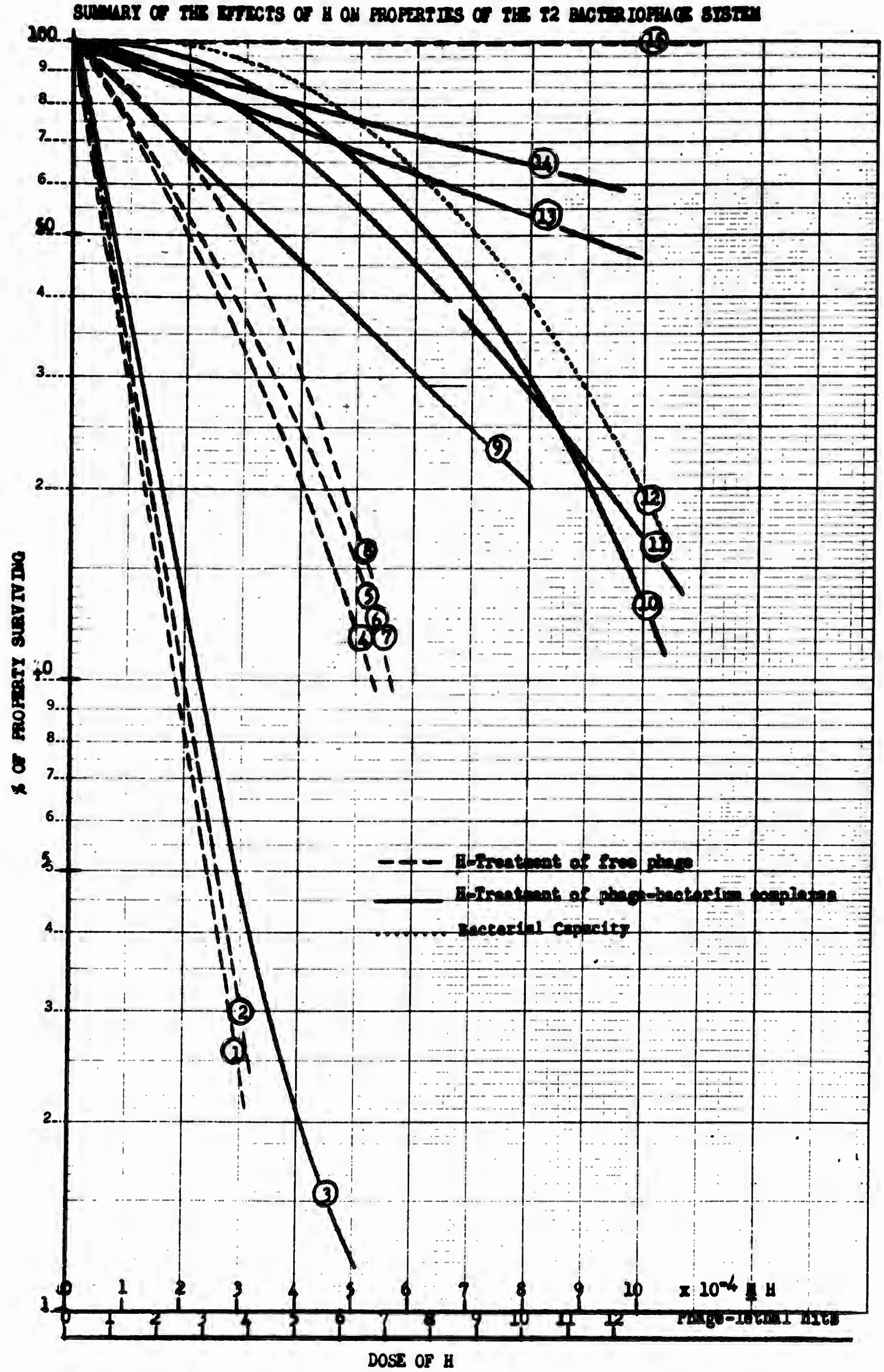


TABLE I

DISTRIBUTION OF HS^{35} ON PROTEIN AND DNA OF T2 BACTERIOPHAGE

Sample	Initial $[HS^{35}]$	Prehydrolyzed $[HS^{35}]$	P/Po	Number of Phage-Lethal Hits (Extracellular)	C.P.M. Per Phage	Molecules* of HS^{35} Bound Per Phage	Molecules of HS^{35} Bound Per Phage Per Phage-Lethal-Hit
10^{12} T2	0	5×10^{-4}	1.0	0	3.6×10^{-11}	3.4	
10^{12} T2	2×10^{-4}	3×10^{-4}	0.018	4.3	3.2×10^{-9}	306	71
10^{12} T2	5×10^{-4}	0	7.9×10^{-5}	10.4	6.6×10^{-9}	630	61

S^{35} Bound to Phage Protein				S^{35} Bound to Phage DNA			
C.P.M. Per Phage After Shocking and DNAase (TCA-Insoluble)	C.P.M. Per Phage After Heating TCA Mixture 95°C (TCA-Insoluble)	% of C.P.M. on Phage Protein	Molecules HS^{35} Bound to Protein Per Phage Per Phage-Lethal Hit	C.P.M. Per Phage After Shocking and DNAase (TCA-Soluble)	C.P.M. Per Phage After Heating TCA Mixture at 95°C (TCA-Soluble	% of C.P.M. on Phage DNA	Molecules HS^{35} Bound To DNA Per Phage Per Phage-Lethal-Hit
3.3×10^{-11}	3.8×10^{-11}			5.1×10^{-12}	2.0×10^{-12}		
1.3×10^{-9}	1.4×10^{-9}	41-44	29-31	2.0×10^{-9}	2.1×10^{-9}	62-65	44-46
2.7×10^{-9}	2.8×10^{-9}	43-45	26-28	3.8×10^{-9}	4.2×10^{-9}	60-67	37-41

* 1.06×10^{-11} c.p.m. = 1 molecule HS^{35}.

TABLE II

EFFECT OF H ON THE DNA SYNTHESIS CATALYZED BY E. COLI POLYMERASE

Experiment[1]	DNA Synthesized in 30 min (% of control)
1. Complete system (control)	100.0
2. Complete system less dTTP	5.9
3. Complete syste, H-treated dATP	81.2[2]
4. Complete system, H-treated dGTP	21.7[2]
5. Complete system, H-treated dTTP	112.0[2]
6. Complete system, H-treated dCTP	124.0[2]
7. Complete system, H-treated polymerase	95.3,[3] 91.0,[4] 82[2]
8. Complete system, H-treated primer[5]	16.8[2]

1/ Complete system (0.3 ml final volume) contained 20 μmoles glycine buffer, pH 9.2; 2 μmoles $MgCl_2$; 0.3 μmoles 2-mercaptoethanol; 10 μg calf thymus DNA (highly polymerized, obtained through the courtesy of Dr. R. M. Herriott); 5 mμmoles dATP; 5 mμmoles dTTP; 10 mμmoles of dGTP; 5 mμmoles of $dCP^{32}PP$ (2×10^6 c.p.m. per μmole); 0.06 units partially purified *E. coli* polymerase.

Initial H

2/ 1×10^{-3} M

3/ 2×10^{-4} M

4/ 5×10^{-4} M

5/ Highly polymerized calf thymus DNA

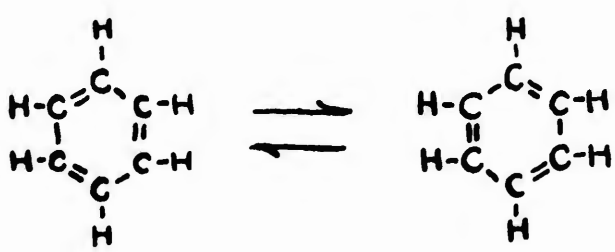
EFFECT OF IONIZING RADIATION ON PYRIDINE

C. PEARCE
U. S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY
FORT MONMOUTH, NEW JERSEY

INTRODUCTION

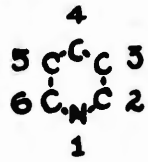
Protection against radiation damage is frequently observed in systems containing benzene or a benzene derivative. The cyclohexane-benzene system (1), where hydrogen and polymer formation from cyclohexane is reduced by the presence of benzene, is well known. Styrene and phenol polymers are relatively stable to radiation. Recently the use of benzene additives to stabilize plastics to radiation damage has been described (2). As a medium for the transfer of excitation energy due to irradiation of solutions, benzene and its derivatives are widely used in liquid scintillators. In fact, the scintillating compound itself may be a benzene derivative such as anthracene or terphenyl. Finally, these benzene compounds are characterized by a relatively low amount of change due to ionizing radiation, compared to other classes of chemical compounds.

Explanations of the resistance of benzene compounds to radiation-induced change, and of the radiation protection and energy transfer mechanisms involve their aromatic structure of six resonating π electrons, i.e.

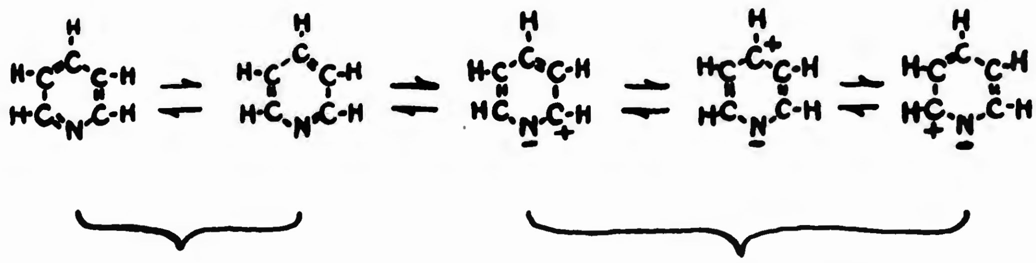


In the structures shown above the π electrons are situated in the double bonds, shown in different positions in the two rings. The π electrons are not confined to any one position in the ring, so the

true benzene structure is an intermediate between the structures above. The benzene ring is considered a hybrid of these two structures and is not a mixture of two individual compounds. Like benzene, pyridine is described as a hybrid of resonating π electrons. But polar contributions are also present because of the presence of a nitrogen atom in the ring. The nitrogen attracts electrons from the 2 and 4 positions, using the numbering system



The resonating structures are:



Aromatic contribution

Polar contribution

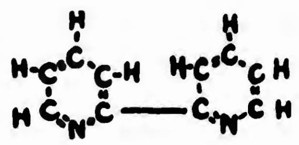
Thus pyridine may have both aromatic and polar characteristics. We were interested to learn which, if either, of these characteristics would be shown by pyridine when irradiated. Should pyridine behave as an aromatic, the radiation protection and energy transfer characteristics noted for benzene could appear. Unlike benzene, pyridine is compatible with both organic and inorganic compounds. This means a wider range of systems may show similar radiation protection and energy transfer characteristics. New energy transfer or conduction and storage systems may be feasible. Further, knowledge of the effect of radiation on pyridine, a unit present in some enzymes and other naturally occurring products, would be of biological importance.

EXPERIMENTATION

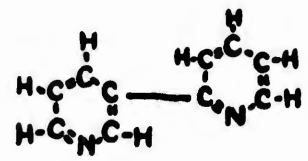
<u>Materials</u>. Pyridine: the Fisher Certified reagent was dried and distilled from calcium hydride, using a 50-plate bubble cap vacuum jacketed column (Bruun); the index of refraction was $n_D^{23.5^\circ C} = 1.5084$.

Bipyridines: 2,2′-bipyridine was the Fisher

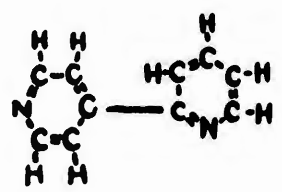
C. P. reagent. Other bipyridines were synthesized and purified as follows: 2,4'- and 4,4'- by the action of sodium upon pyridine(3); 2,3'- and 3,3'- from the corresponding phenanthrolines(4). Synthesis of 3,4'-bipyridine has not been accomplished yet, since a gas chromatography column has not been found to completely resolve the bipyridines. The respective structural formulae are:



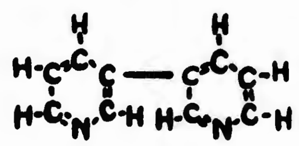
2,2'-Bipyridine



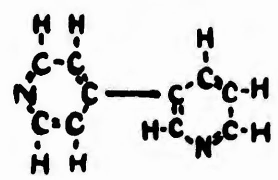
2,3'-Bipyridine



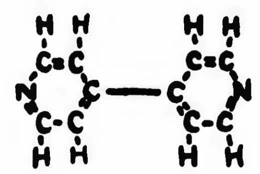
2,4'-Bipyridine



3,3'-Bipyridine



3,4'-Bipyridine



4,4'-Bipyridine

Calibration solutions for gas chromatograms: pyridine solutions of the five available bipyridines, 0.13 M in each bipyridine; both individual and composite solutions were prepared.

<u>Irradiation</u>. Exposures (about 50 hours) were made using a 120-curie Cs^{137} source (0.663-Mev gamma radiation). Dose rates were measured with the ferrous sulfate (Fricke) dosimeter using $G_{(Fe^{+++})} = 15.6$ (5); the value determined, 3.4×10^{20} electron volts/liter/hour or 5410 roentgens/hour, was corrected for the electron

density of pyridine, becoming 4.9×10^{20} electron volts/liter/hour. Ferric concentrations were determined spectrophotometrically at 3040 Å.

Procedure. Using an all-glass high vacuum system, capable of 10^{-6} to 10^{-7} mm range Hg pressure, pure dry pyridine was degassed by a freeze-pump-thaw cycle repeated three times. To reduce contamination from vacuum stopcock grease, about 25 ml were distilled into a trap, thence to a second trap (a calibrated tube), and finally a measured volume was distilled into the irradiation cell, and the cell sealed off.

Following irradiation, the cell was attached to the vacuum line, opened, and the contents degassed as before. A Toepler pump was used to transfer the gaseous products thru two U-traps immersed first in dry-ice snow and then in liquid nitrogen. The non-condensable gas at each temperature was collected in the gas thimbles of the Blacet-Leighton apparatus.

The excess pyridine was distilled slowly into a reservoir. A small amount of pyridine was slowly distilled back to wash the residue down the walls, before the tube was cut off and the contents analyzed.

Analysis. Gas chromatography (or GC): non-gaseous radiolysis products were determined, in pyridine solution, on a Perkin-Elmer Vapor Fractometer Model 154 C. The 12-foot stainless steel column was packed with 0.25% Carbowax 1500 (polyethylene glycol) on glass beads (60/80 mesh). The column was maintained at 141 ± 1°C, the flow rate of the helium carrier gas was 25.4 ml/min. This column gave much better resolution of bipyridines, than any other prepared. Even so, as noted below, the resolution was far from optimum and the retention times were extremely sensitive to temperature, solute concentration, and sample size.

Gas analysis: gaseous radiolysis products were analyzed on a Blacet-Leighton (Arthur H. Thomas Co. modification) micro gas analysis apparatus, according to the procedures described by Scott(6). Analysis were conducted for H_2, C_2H_2, C_2H_4, CH_4, and air (O_2, and N_2 by difference).

Spectroscopy: ultraviolet absorption spectra of the non-gaseous radiolysis products were determined on a Cary Model 11 Recording Spectrophotometer, using 1 cm silica cells. IR spectra of these products were taken on a Perkin-Elmer Model 21 Double Beam Infrared Spectrophotometer equipped with a sodium chloride prism. Ferric ion concentrations were determined on a Beckman DUV Spectrophotometer using 1 cm silica cells.

RESULTS AND DISCUSSION

The major product of the irradiation of pyridine is about

2 mg. of greenish yellow "varnish" and dark brown oily residue. The amount of gas formed was negligible. Small amounts (less than 150 mm^3) of gas, non-condensable at liquid nitrogen temperatures, were collected in the thimbles. Upon analysis however, these samples proved to be air (20% O_2 + 80% inert gas, believed to be nitrogen) which may have been trapped in the Toepler pump outlet and carried out into the gas thimble.

A typical gas chromatography trace of the non-gaseous fraction is shown in Figure 1. Comparison of the traces with the calibration curve (Figure 2) identified the first peak as 2,2′-bipyridine. The second peak could not be identified as easily due to the lack of resolution for the remaining bipyridines.

Some interpretation of the GC traces is possible however. One does find a "clumping" of bipyridines in the calibration curve when compared to traces for the individual bipyridines. The retention times for the bipyridines, after normalization to the 2,2′-bipyridine peak as zero time are:

Bipyridine	Retention time, seconds corrected to 2,2′-bipyridine	Bipyridine boiling point, °C
2,3′-	795 ± 30	296
2,4′-	880 ± 33	297
3,3′-	1083 ± 47	301
4,4′-	1115 ± 30	305

The corresponding times for the second and third peaks of the calibration solution are 753 ± 53 and 936 ± 60 seconds. The lowering of retention times is due likely to the increased (5 times) solute concentration in the calibration solution as compared to the individual bipyridine solutions. The appearance of each isomer in the individual solutions was in order of its boiling point. Had 3,4′-bipyridine been present, it would have appeared later than the 2,4′-isomer(7) accordingly. We have assigned the peaks in the calibration solution as 753 seconds representing 2,3′- and 2,4′-bipyridine and 936 seconds the 3,3′- and 4,4′-compounds, after considering the respective boiling points and observed retention times.

The GC traces for the non-gaseous radiolysis product show a second peak at 665 seconds, normalized to the 2,2′-bipyridine peak. We believe this corresponds to the 753 second peak of the calibration curve. Although the concentration of solute was less than that of the calibration solution, the sample volume was two to threefold greater in order to put the largest amount of radiolysis product possible on the column. This larger sample volume may account for the reduced retention time.

Further evidence for the presence of 2,2′- and one or both of 2,3′- and 2,4′-bipyridine in the residue is found in the ultraviolet absorption spectrum of the residue dissolved in 0.02 M $NH_4OH_{(aq.)}$ (Figure 3). Figure 4 shows the corresponding spectrum of equal amounts of 2,2′-, 2,3′-, 2,4′-, 3,3′-, and 4,4′- in 0.02 M $NH_4OH_{(aq.)}$ solution. Absorption maxima determined for the individual bipyridines(8), indicated on both figures, are:

Bipyridine	λ,Å	λ,Å
2,2′-	2800	2330
2,3′-	2750	2370
2,4′-	2730	2380
3,3′-	2690	2390
3,4′-	Inflection point at 2680	2410
4,4′-	Inflection point at 2700	2390

Difference in absorption or optical density are due to concentration differences. Most interesting is the position of the peak in the 2700-2800 Å region, which, for the solution of five bipyridines, appears midway (2730 Å) between the peaks given for the individual isomers. In Figure 3, the curve for the radiolysis product peaks at 2780 Å, midway between the 2,2′-, 2,3′-, and 2,4′- peaks. The bipyridines contributing to these peaks in Figure 3 must be among the 2,2′-, 2,3′-, and 2,4′-bipyridines. Infrared absorption spectra conducted on the pyridine solutions of the residue and on individual bipyridines did not resolve the identity of the residue further. Accordingly, the bipyridines are identified as the 2,2′- and 2,x′-isomers.

The yields of non-gaseous products were also determined from the GC calibration curve for the bipyridine solution. The amount of 2,2′-bipyridine was calculated from the peak area of the corresponding peak of the GC curve. The peak height of the 2,x′-bipyridine peak was correlated with the peak height of the 2,3′- + 2,4′-bipyridine peak on the calibration curve to determine the amount. Yields of product were determined as that formed per 100 electron volt energy input, and were calculated according to the formula

$$G = \frac{100 \times 6.02 \times 10^{23} \times M}{E \times V}$$

where $G_{product}$ = yield of product/100 ev

M = moles product

E = irradiation dose, electron volts/liter

and V = volume of pyridine irradiated, liters

The yields so calculated were:

$$G_{2,2'\text{-bipyridine}} = 0.11$$

$$G_{2,x'\text{-bipyridine}} = 0.20$$

$$G_{\text{bipyridine}} = 0.31$$

To determine if the results above represented the total non-gaseous radiolysis products, the weight of the product before GC analysis, was determined and compared with that calculated from the GC trace. Only about one-half the product is accounted for on the GC traces, so other compounds must be also present.

CONCLUSION

The gamma irradiation of liquid, oxygen-free and dried pyridine resulted in extremely low yields of product. By comparison, benzene, a chemical compound considered relatively stable to radiation, gave $G_{(H_2)} = 0.03$ and $G_{(\text{polymer})} = 0.98$ (1) (9). The gas yield from pyridine was undetectable by our methods. Had the H_2 yield been as large as that from benzene, about 7 mm^3 would have been formed which would have been measurable.

The polymer yield from benzene compares to our total bipyridine yield of 0.31 plus that of the non-gaseous products not yet identified. Hydrogenated dimer products ($C_{10}H_{10}N_2$, $C_{10}H_{12}N_2$, $C_{10}H_{14}N_2$, etc.) and even trimers may be present. These dimer products would not differ greatly in molecular weight from the bipyridines $C_{10}H_8N_2$), so that the yield (note weight measurements) should be somewhat similar. The asymmetry due to the N-atom of the pyridine ring, results in a great number of dimer (and trimer) isomers. The concentration of any one of these products would not likely be high enough to be detectable on the GC traces. Combustion analyses to determine C-H-N ratios, as well as molecular weight determinations, are necessary to clarify the remainder of the non-gaseous product. However, the $G_{(\text{pyridine} \rightarrow \text{polymer})}$ according to our bipyridine and weight measurements would appear to be less than 1.

Aromatic compounds are characterized by low radiolysis yields, so that pyridine with total product yields no greater than that for benzene, shows aromatic behavior when irradiated. The nature of the non-gaseous product also conforms to the aromatic pattern, i.e. bipyridines compare to biphenyl (formed from benzene). A contribution from the polar nature of pyridine is found in the identity of the bipyridines formed (2,2'-; both or either of 2,3'-, 2,4'-). Quantum mechanical treatment of the pyridine resonance structure suggests the 2 position is most favored for radical attack and the 3 and 4 positions are next most favored equally(10). Further, though the type and total yields of product are comparable for pyridine and benzene, the distribution of products is different.

With benzene $G_{(biphenyl)} = 0.065$ (9) whereas for pyridine $G_{(bipyridine)} = 0.31$. One infers that hydrogenated products are formed in much smaller amounts for pyridine.

Consistent with the results, a radiolysis mechanism based upon the formation of free radicals has been formulated. This mechanism must account for the bipyridine formation, the absence of H_2 gas, and the formation of non-gaseous products so far unidentified, but perhaps hydrogenated isomers.

$$C_5H_5N \rightsquigarrow C_5H_5N^*$$
excited pyridine molecule

$$C_5H_5N^* \rightarrow C_5H_4N^\bullet + H^\bullet$$
free radical formation

Radical recombination

$$C_5H_4N^\bullet + C_5H_4N^\bullet \rightarrow C_{10}H_8N_2$$
bipyridines

may occur as well as reaction of the radicals with the solvent:

$$C_5H_5N + C_5H_4N^\bullet \rightarrow C_{10}H_8N_2 + H^\bullet$$
bipyridines

Interaction of $H^\bullet$ with other species present, i.e. C_5H_5N, $C_5H_4N^\bullet$, and $C_{10}H_8N_2$, could result in hydrogenated products, or in species which could react further to form such products. The recombination of $H^\bullet$ to form H_2 is not presented since the gas yield was undetectable. The formation of H_2 during benzene radiolysis is thought to come from non-radical processes(11), which would be consistent with our results and mechanisms.

Our results show pyridine reacts chiefly as an aromatic compound when irradiated. Important applications of this work may be possible in energy transfer media, in radiation protection, in biological studies, in reactor coolants, and in the creation of new radiation resistant materials.

ACKNOWLEDGMENT

The author acknowledges the assistance of Mr. Joseph E. Ellison, Jr., Chemist, Institute for Exploratory Research, in performing the experimental work.

REFERENCES

1. J. Lamborn and A. Swallow, J. Phys. Chem. 65, 920 (1961).

2. D. Gardner and L. Epstein, J. Chem. Physics 34, 1653 (1961).

3. C. Smith, J. Am. Chem. Soc. 46, 414 (1924).

4. C. Smith, J. Am. Chem. Soc. 52, 397 (1930).

5. C. Hochanadel and J. Ghormley, J. Chem. Phys. 21, 880 (1953)

6. W. Scott, "Standard Methods of Chemical Analysis," Fifth Edition, Van Nostrand Co., Inc., New York, N.Y., pp 2537-2545, 1939.

7. F. Kuffner and F. Straberger, Monatshefte 88, 793, (1957).

8. P. Krumholz, J. Am. Chem. Soc. 73, 3487 (1951).

9. T. Gäumann and R. Schuler, J. Phys. Chem. 65, 703 (1961).

10. G. Wheland, "Resonance in Organic Chemistry," Wiley & Sons, Inc., New York, N.Y., 1955.

11. J. Burr and J. Scarborough, J. Phys. Chem. 64, 1367 (1960).

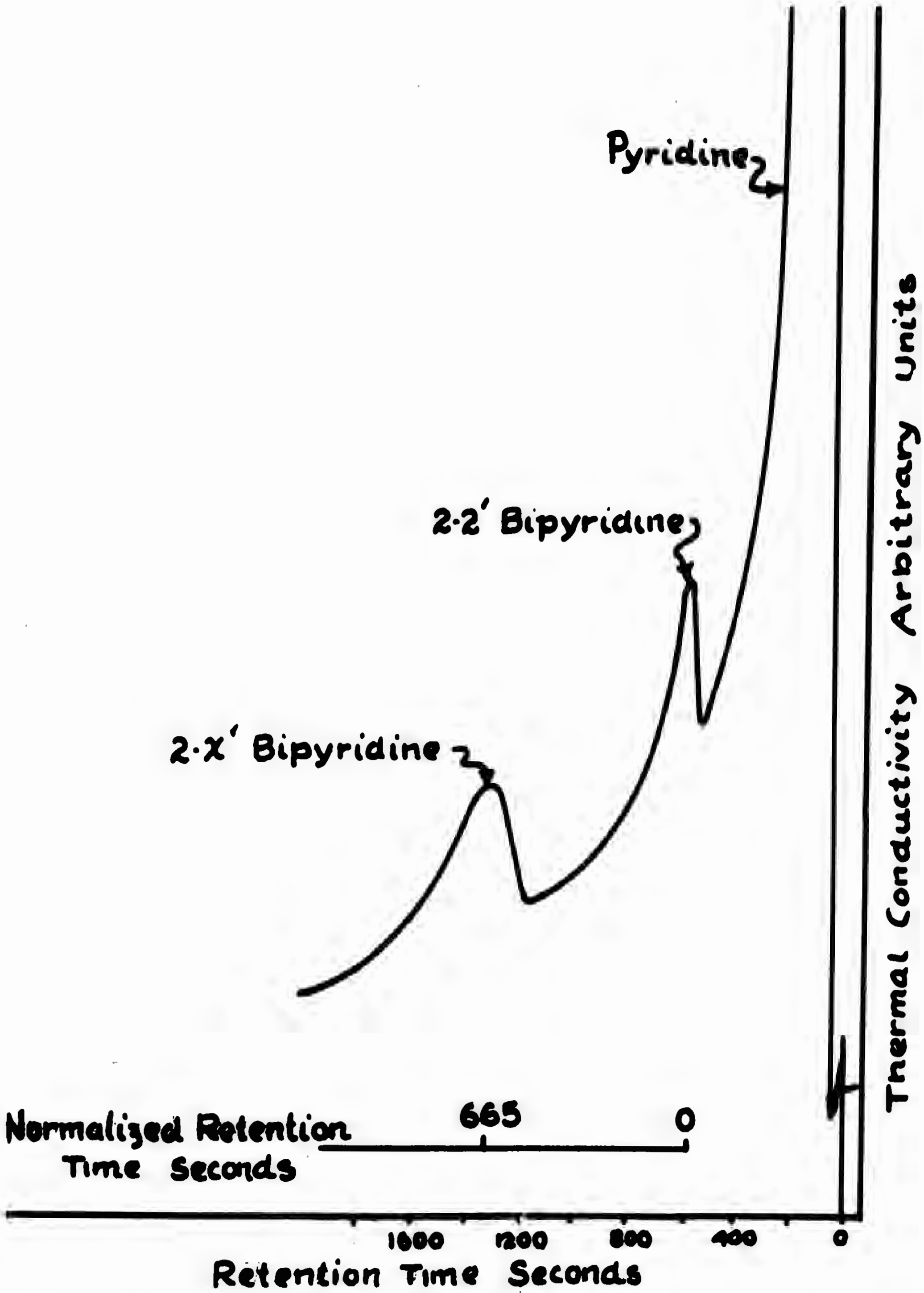


FIGURE 1
Gas Chromatogram of Non-Gaseous Products from Irradiated Pyridine

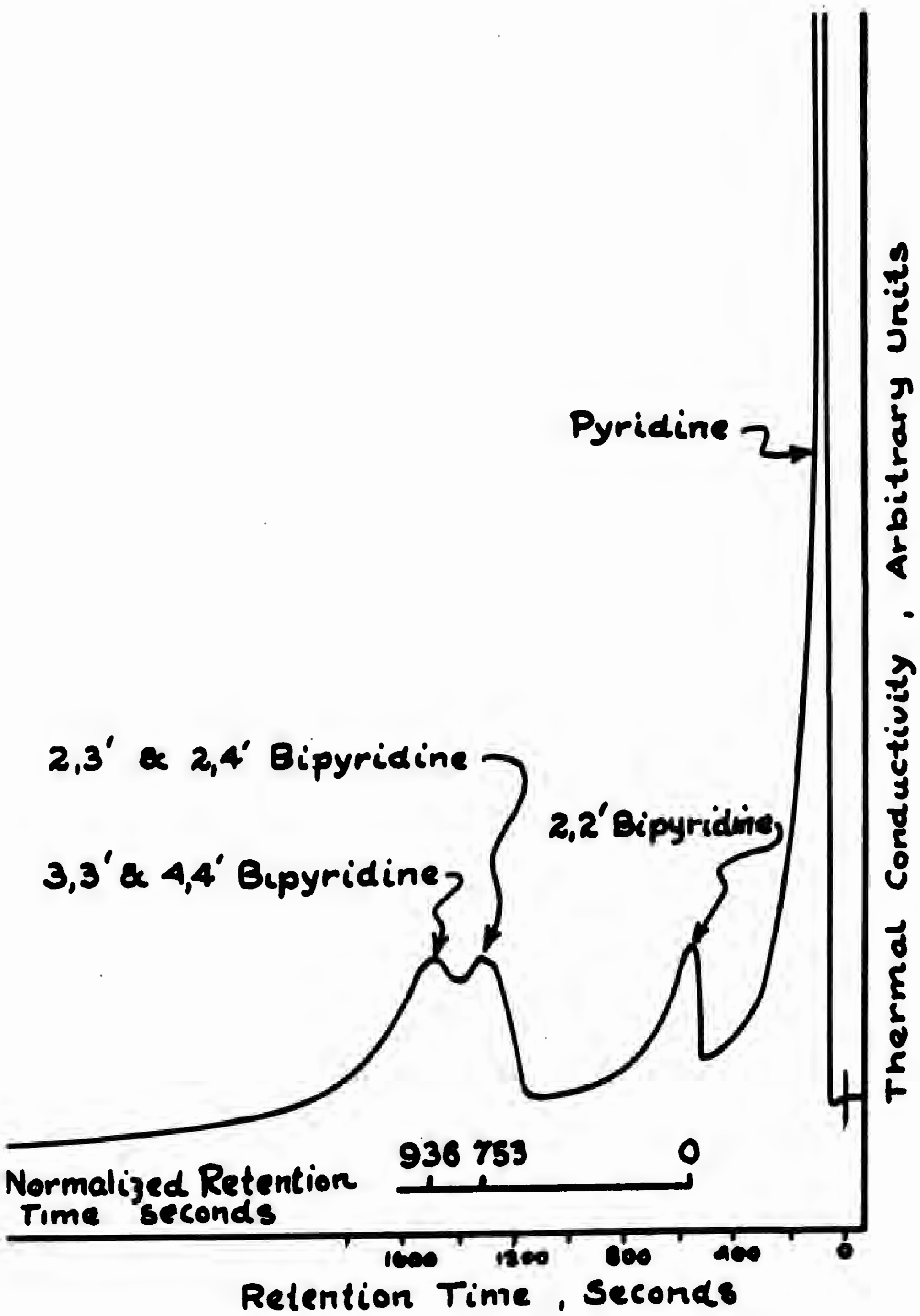


Figure 2

Gas Chromatogram of Bipyridine Calibration Solution

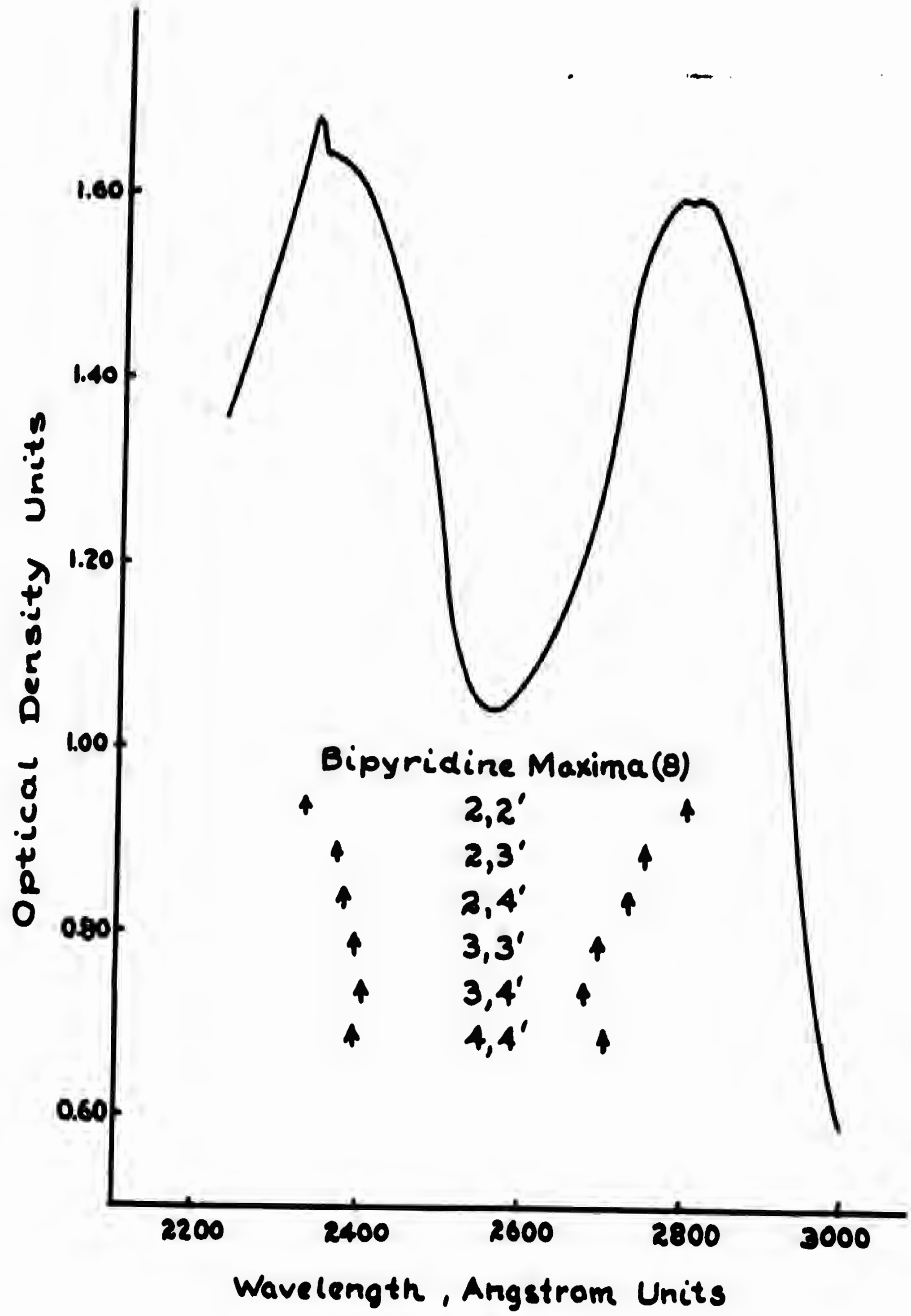


Figure 3

Ultraviolet Absorption Spectrum of Non-Gaseous Products from Irradiated Pyridine

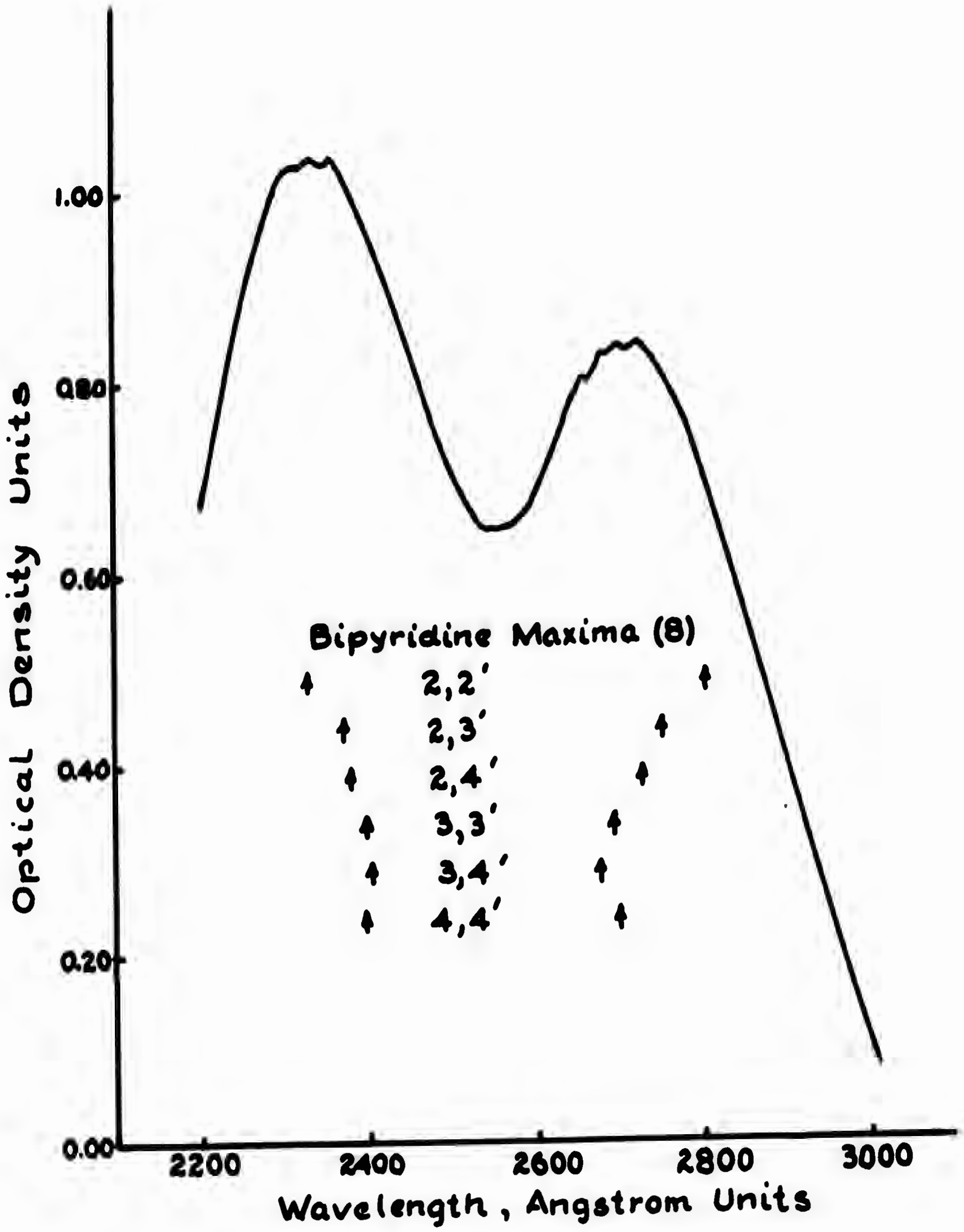


Figure 4

Ultraviolet Absorption Spectrum of Solution of 2,2' ; 2,3' ; 2,4'; 3,3'; 4,4'; 3,4'; Bipyridines

PARTIALLY FLUORINATED ALIPHATIC COMPOUNDS BY REDUCTIVE DESULFURIZATION OF SUBSTITUTED THIOPHENE

S. Portnoy and H. Gisser
Pitman-Dunn Laboratories
U. S. Army Ordnance Corps
Frankford Arsenal, Philadelphia, Pa.

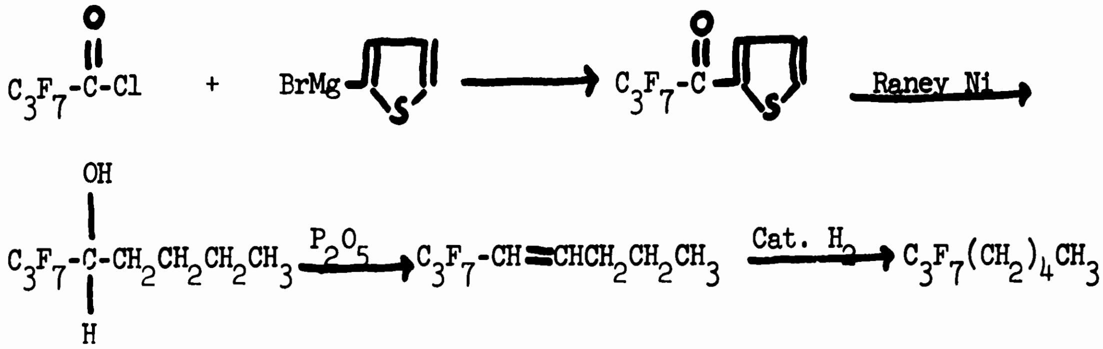
The reductive desulfurization of substituted thiophenes has been proposed as a synthetic route for the preparation of long-chain acids, ketones, alcohols and hydrocarbons.[1] The preparation of a straight-chain saturated hydrocarbon was accomplished by the condensation of thiophene and an acid in the presence of phosphorus pentoxide, Clemmensen reduction of the resulting ketone, and finally, desulfurization with Raney nickel. Lengthening of the chain can be accomplished by further reaction in the 5-position of thiophene. Various branched-chain compounds can be prepared by the use of suitably β-substituted thiophenes.

The purpose of our investigation was to determine whether this synthetic technique could be utilized for the preparation of partially fluorinated compounds and thus provide a method for preparation of heat-stable fluids. As described previously,[2] the preparation of 2-heptafluorobutyrylthiophene was achieved by a reverse Grignard technique. The attempted condensation of heptafluorobutyric acid and thiophene in the presence of phosphorus pentoxide, resulted only in the conversion of the fluorinated acid to the acid anhydride. Other methods were equally unsuccessful.

Reduction of 2-heptafluorobutyrylthiophene was unsuccessful via Clemmensen or a Wolff-Kishner reaction using in the latter either 85% hydrazine hydrate or semicarbazide hydrochloride. Reductive desulfurization with Raney nickel resulted in the isolation of the alcohol 1,1,1,2,2,3,3-heptafluoro-4-octanol. The partially fluorinated olefin and hydrocarbon were then prepared as shown by the following over-all scheme.

(1) H. Wynberg and A. Logothetis, J. Am. Chem. Soc., 78, 1958 (1956)

(2) S. Portnoy and H. Gisser, J. Org. Chem., 22, 1752 (1957)



In the reductive desulfurization reaction, a small amount of 1,1,1,2,2,3,3-heptafluoro-4-octanone was also isolated. Attempted reduction of this ketone by the Clemmensen or Wolff-Kishner reactions was unsuccessful. However, the ketone was easily converted in high yield to the alcohol by treatment with $LiAlH_4$.

EXPERIMENTAL[3]

2-Heptafluorobutyrylthiophene. An ether solution of 1 mole of thiophene magnesium bromide was added slowly to an ether solution of 1.2 moles of heptafluorobutyryl chloride and worked up in a manner previously described.[2] Distillation gave a 32% yield, b.p. 91.5-92.1° at 32 mm., n_D^{20} 1.43186.

Anal. Calcd. for $C_8H_3OF_7S$: C, 34.29; H, 1.08; F, 47.47. Found: C, 34.13; H, 1.25; F, 46.85.

The 2,4-dinitrophenylhydrazone, recrystallized from an alcohol-water mixture, melted at 90.2-90.8°.

1,1,1,2,2,3,3-Heptafluoro-4-octanol. 2-Heptafluorobutyrylthiophene (25 g., 0.09 mole), 250 g. Raney nickel[4] and 900 ml. of 95% alcohol were stirred and refluxed for 16 hr. The alcohol was decanted and the nickel was washed with alcohol and ether. The solvents were distilled off, the remaining liquid dried overnight over anhydrous Na_2SO_4 and filtered.

(3) All temperature readings are uncorrected. The analyses were carried out by Wyanel Laboratories, Philadelphia, Penna. and by Schwarzkopf Microanalytical Laboratory, Woodside, N. Y.

(4) R. Mozingo, Organic Synthesis, Coll. Vol. III, E. C. Horning ed., J. Wiley and Sons, Inc., New York, 1955, p. 181

Fractionation gave 1,1,1,2,2,3,3-heptafluoro-4-octanol (9.5 g., 41%) b.p. 148.1-148.6°, n_D^{25} 1,34445.

Anal. Calcd. for $C_8H_{11}OF_7$: C, 37.51; H, 4.33; F, 51.92. Found: C, 38.31; H, 4.64; F, 52.65.

There was also isolated a small amount of 1,1,1,2,2,3,3-heptafluoro-4-octanone (1.0 g., 4%), b.p. 119.7-121.3°, n_D^{25} 1.326, semicarbazone 92.4-95.3°.These values correspond to that found in the literature.[5]

1,1,1,2,2,3,3-Heptafluoro-4-octene. Phosphorus pentoxide (56.8 g., 0.4 mole) was placed in a 500 ml. flask and to this was added 1,1,1,2,2,3,3-heptafluoro-4-octanol (102.5 g., 0.4 mole). The mixture was distilled and the material from 59-124° collected. The distillate was washed with 5% $NaHCO_3$ until slightly alkaline and then washed twice with distilled water. The combined washings were extracted once with ether. The ether-distillate solution was dried overnight over anhydrous Na_2SO_4, filtered, and stripped of solvent. Fractionation through a 3 in. column with 1/8 in. helices gave 1,1,1,2,2,3,3-heptafluoro-4-octene (25.4 g., 27%), b.p. 105.3-106.2°, n_D^{25} 1,32980. Infrared spectra (max. at 5.98μ) confirmed the olefin.

(5) K. T. Dishart and R. Levine, J. Am. Chem. Soc., 78, 2268 (1956)

DYNAMIC SIMULATOR

F. PRADKO
U.S. ARMY ORDNANCE TANK-AUTOMOTIVE COMMAND
DETROIT, MICHIGAN

INTRODUCTION

Recently much effort has been devoted to the study of shock and vibration effects upon military vehicles. Particular attention has centered in two areas:

I. Study of vehicle suspension systems.

II. Vehicle reaction to gun recoil forces.

This document presents the results of research in each of these fields.

Program activity involved various engineering sciences, including mathematics, computer techniques and motion simulation. Within the framework of each element, specific knowhow was developed and joined together. The resulting product was a new engineering capability, now in use at the Ordnance Tank-Automotive Command in Detroit.

The scope of suspension work considered a full range of environmental conditions that were deemed possible and worthy of study. The operational modes reviewed began with highway travel and extended to plowing over rough terrain, as shown in FIG. 1.

FIG.1. Truck, 1 Ton, 8x8, XM-384

The study of vehicle firing stability treated design criteria in a similar unrestricted fashion. Chassis dynamics were assumed to generate from either recoilless rifle action or large caliber cannon shots as exhibited in FIG. 2.

FIG.2. Self Propelled Artillery, T-236

OBJECTIVE

This research was fostered by OTAC to move Army vehicle design capability forward to bolster the 1961 state of knowhow and provide necessary preparedness for new technical requirements.

The underlying interest that launched the program emerged from the disturbing awareness that the value of military equipment can be a very temporary condition, since design supremacy is continually challenged by foreign competition. First place is a precarious position that is only maintained by moving ahead. Progress, however, depends upon knowhow for knowhow is a blend of two items, theory and experience. Theory provides the technical understanding to do a job and experience develops the mechanical ability to bring about the desired result.

The function of theory is not simply a long hair study of fundamentals. In the proper mode the primary attribute of theoretical activity is development of systematic thinking, to explore new subjects profitably. This contribution is vital, for insight is developed to release new ideas keeping experience flexible and free from decay. Thus, the science describing military vehicle development must remain in the area of experience; however, it should also be periodically refreshed.

The subject research, (1) Suspension Design and (2) Firing Stability Analysis, are two elements of vehicle design that are perennials, both may be described as "tough nuts" for each is conspicuous by continuous appearance in new design programs. Individually they present difficult problems to vehicle engineers with many interlaced variables. The magnitude of each task has produced experts and design specialists. These people, however, are not medicine men who are able to consistently generate successful answers. They are merely endowed with creativity, producing practical approaches based on past experience or new research. Consequently, knowledge of suspension and stability analysis needs new technical substance to make the job of arriving at correct solutions easier for design practitioners. This required technical input is not readily filled. Growth is not automatic. Without a pointed stimulus, engineering areas such as these

normally exhibit slow expansion.

Accordingly, Research management at the Ordnance Tank-Automotive Command felt that suspension and firing stability dynamic studies should not remain in this class, for new capability was continually in demand. Techniques were needed to predict and simulate the performance of vehicle dynamics, both analytically and in the physical form. Two method representation was felt to be necessary for maximum engineering usefulness: An analytical option for opportune technical analysis and physical simulation to pretest designs before release from the drafting board.

APPROACH

The directed attainment of these objectives was channeled into a three phase pattern:

First was the usual problem analysis activity to clarify problem details, so that the solution would not fail to fill the bill.

Next came the search and research phase to establish facts and technical evidence upon which to base solutions.

Then came exercise and application of the resulting techniques.

Suspension Analysis:

The three point approach was initially applied to the Suspension System work. Suspension systems of military vehicles were considered to have design importance for one reason, mainly to dissipate, without objectionable reaction, the destructive kinetic energy forces that develop between a moving vehicle and the road surface. The reactions studied were impact forces causing structure oscillation and component vibration.

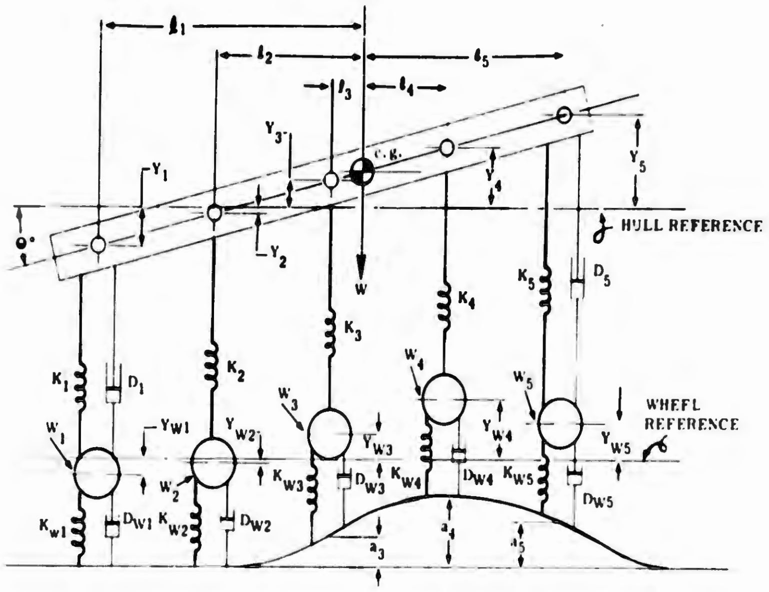


FIG.3. Block Diagram Vehicle Suspension

Vehicle motions were resolved into pitch, bounce and roll components. Equations were derived from typical schematic diagrams of vehicle mass distribution, spring rates and damping characteristics, as follows:

<u>Hull Bounce Acceleration ($\ddot{y}$)</u>:

Vertical acceleration of the vehicle hull is influenced by spring forces, shock absorber damping forces and the sprung weight of the vehicle.

$$\ddot{y} = -\frac{k_1}{M_o}(y_1 - y_{w1}) - \frac{k_2}{M_o}(y_2 - y_{w2}) - \frac{k_3}{M_o}(y_3 - y_{w3}) - \frac{k_4}{M_o}(y_4 - y_{w4}) - \frac{k_5}{M_o}(y_5 - y_{w5}) - \frac{D_1}{M_o}(\dot{y}_1 - \dot{y}_{w1}) - \frac{D_5}{M_o}(\dot{y}_5 - \dot{y}_{w5}) + g$$

Hull Pitch Acceleration ($\ddot{\theta}$):

The angular pitch acceleration is proportional to the torques about the c.g. due to spring forces, damping forces and the polar moment of inertia of the sprung weight.

$$\ddot{\theta} = - \frac{k_1 l_1}{J_o} (y_1 - y_{w1}) - \frac{k_2 l_2}{J_o} (y_2 - y_{w2}) - \frac{k_3 l_3}{J_o} (y_3 - y_{w3}) + \frac{k_4 l_4}{J_o} (y_4 - y_{w4}) + \frac{k_5 l_5}{J_o} (y_5 - y_{w5}) - \frac{D_1 l_1}{J_o} (\dot{y}_1 - \dot{y}_{w1}) + \frac{D_5 l_5}{J_o} (\dot{y}_5 - \dot{y}_{w5})$$

Bounce Acceleration of Road Wheels ($\ddot{y}_w$):

Vertical acceleration is dependent upon the spring and damping force of the tire, the spring and shock absorber force of the suspension and the weight of the wheels.

$$\ddot{y}_{w1} = - \frac{k_{w1}}{M_{w1}} (y_{w1} - a_1) - \frac{D_{w1}}{M_{w1}} (\dot{y}_{w1} - \dot{a}_1) + \frac{k_1}{M_{w1}} (y_1 - y_{w1}) + \frac{D_1}{M_{w1}} (\dot{y}_1 - \dot{y}_{w1}) + g$$

Bounce Displacement of the Hull (y):

The vertical displacement of the hull over each wheel is equal to the displacement of the center of gravity plus the horizontal distance to the center of gravity times the sine of the pitch angle.

$$y_1 = y_o + l_1 \sin \theta$$

Bounce Velocity of the Hull ($\dot{y}$):

The vertical velocity of the hull over each wheel is equal to the vertical velocity of the center of gravity plus the horizontal distance from the center of

gravity times the sine of the pitch angular velocity.

$$\dot{y}_1 = \dot{y}_0 + l \cdot \sin\theta$$

<u>Generation of Road Profile</u>:

The ability to generate a road profile or terrain condition was found to be as necessary for realistic computer simulation studies as writing the equations of behavior.

The technique developed for this purpose is unique, the road data that is presented to each vehicle wheel is generated on a digital computer. The computer stores the road elevation increments, selects what elevation each wheel requires at a particular time, and generates the time between each increment of elevation. Data is converted from the digital form into equivalent analog form by a D-A Converter. One conversion per each vehicle wheel is necessary. This information then serves as an input to an Analog computer suspension system simulation.

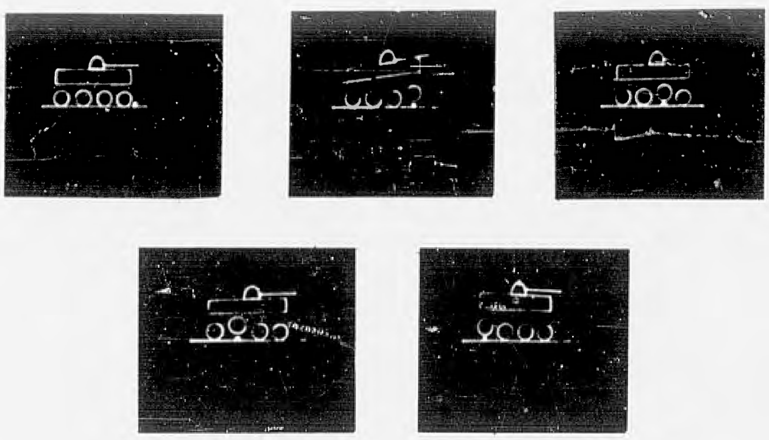


FIG.4. Generation of Road Profiles

Example, FIG.5. illustrates the systems flexibility to construct different roads or terrain contours. OTAC's road profile library consists of the Perryman random cross country terrain located at the Army Ordnance Proving Grounds in Aberdeen, Maryland; the Aberdeen Belgium Block Course, a Ft. Knox tank cross country trail and any form of sine or square wave profile.

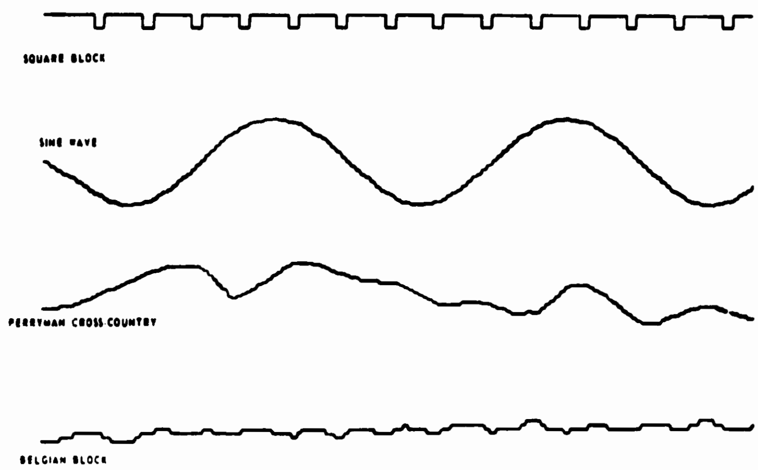


FIG.5. Road Profiles

At this point, the suspension research program produced:

1. Mathematical model of suspension systems.
2. Terrain profile generation techniques.
3. Analog computer simulation procedure.
4. Visual (oscilloscope) display of computer results.

These products, while appearing to present a completely suitable design tool, were not very well received. Designers using the system for a first time invariably brought forth scepticism. Results of suspension studies were not effectively communicated although they were tendered in three forms - graphical, tabulated, and visual. It appeared that these yardsticks were not of the right material to be universally appreciated by program managers, designers, and administrative people. Another gage was in demand, one that would normalize the results so that design merits could be judged first hand, by each individual using the seat of the pants technique.

Simulator:

The answer was this simulator:

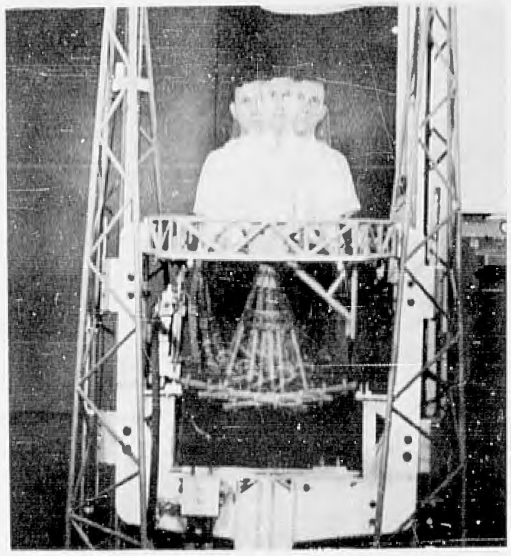


FIG.6. Simulator

Perhaps the most significant claim that can be broadcast for the simulator, at this time, is that it will make possible performance trials of designs prior to building of a design. In some instances it is the only economical approach, considering time and cost, particularly, where a suspension requirement is being investigated by many alternatives.

The simulator is a high amplitude, hydraulically driven and electronically controlled device. It is capable of producing four different degrees of motion, either individually or collectively.

Excursion and Acceleration Rates:

Motion	Max.Tot.Travel	Max.Frequency	Acceleration
Bounce	3 ft	10 cps	2 g's
Roll	40 deg	10 cps	30 radians/sec^2
Pitch	40 deg	10 cps	30 radians/sec^2
Yaw	20 deg	3 cps	15 radians/sec^2

Firing Stability:

The simulator was also used to simulate the dynamics of firing stability. This program began when it

was recognized that military characteristics and design requirements of combat vehicles placed increased emphasis on the need for mounting large caliber weapons on chassis of reduced weight and decreased size. The starting cue was very forceably observed in the M56, 90mm, Self-Propelled vehicle.

FIG.7. Self-Propelled Artillery M-56

In keeping with this indicated trend it was necessary to establish, with greater accuracy, the stability of contemplated designs prior to extensive prototype development.

The characteristic motions of chassis movement were analyzed and equations were developed for each, by application of Newton's laws.

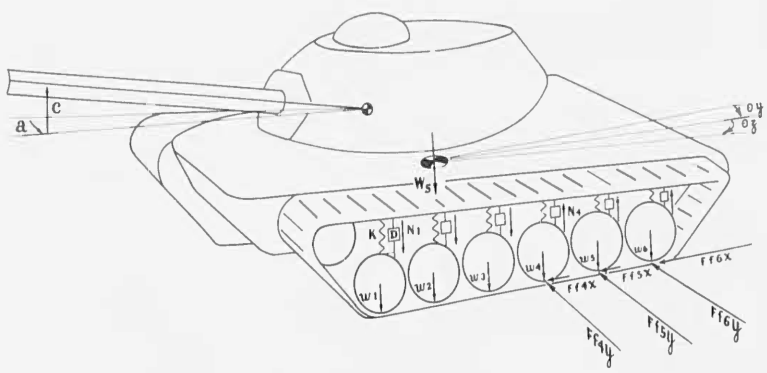


FIG.8. Tank Schematic Diagrams

Equations:

Longitudinal Translation:

Gun Force + Ground Frictional Force = Mass x Acceleration

$$F_{gx} \quad + \quad F_{fx} \quad = \quad m\,(\ddot{x}-\dot{y}\,\dot{\theta}_z + \dot{z}\dot{\theta}y)$$

Lateral Translation:

Gun Force + Ground Frictional Force = Mass x Acceleration

$$F_{gy} \quad + \quad F_{fy} \quad = \quad m\,(\ddot{y}-\dot{z}\,\dot{\theta}_x + \dot{x}\dot{\theta}_z)$$

Bounce:

Gun Force - Sprung Weight + Suspension Forces

$$F_{gz} \quad - \quad W_s c_{zZ} \quad + \quad \sum_{i=1}^{n} N_i$$

= Mass x Acceleration

$$= m\,(\ddot{z}-\dot{x}\,\dot{\theta}_y + \dot{y}\dot{\theta}_x)$$

Roll:

Gun Moment + Ground Frictional Moment + Suspensional

$$-\bar{z}F_{gy} + \bar{y}F_{gz} + (z + z_0)\,F_{fy}c_{yY} \quad + \sum_{i=1}^{n} N_i y_i$$

Moment = Angular Acceleration x Moment of Inertia

$$= \ddot{\theta}_x I_x - \ddot{\theta}_z I_{xz} + (I_z - I_y)\,\dot{\theta}_y\dot{\theta}_z - I_{xz}\dot{\theta}_x\dot{\theta}_y$$

Pitch:

Gun Moment - Ground Frictional Moment - Suspensional

$$-\bar{x}F_{gz} + \bar{z}F_{gx} - (z + z_0)\,F_{fx}c_{xX} \quad - \sum_{i=1}^{n} N_i x_i$$

Moment = Angular Acceleration x Moment of Inertia

$$= \ddot{\theta}_y I_y + \dot{\theta}_x\dot{\theta}_z\,(I_x - I_z) + (\dot{\theta}_x^2 - \dot{\theta}_x^2)\,I_{xz}$$

Yaw:

Gun Moment + Ground Frictional Moments

$$-\bar{y}F_{gx} + \bar{x}F_{gy} + \sum_{i=1}^{n} x_i\,F_{fiy}\,c_{yY} \quad \sum_{i=1}^{n} y_i\,F_{fix}\,c_{xX}$$

= Angular Acceleration x Moment of Inertia

$$= \ddot{\theta}_z I_z - \ddot{\theta}_z I_{xz} + (I_y - I_x)\,\dot{\theta}_x\dot{\theta}_y + I_{xz}\,\dot{\theta}_y\dot{\theta}_z$$

FIG.9. Firing Stability Equations

The derived Stability Equations were for weapon systems free to move in three degrees of angular freedom - roll, pitch and yaw, and three degrees of translational freedom - fore and aft movement, bounce and lateral slip. These equations define the motion of the vehicle, as effected by interrelated factors of gun firing force, gravity, terrain influence, and the resisting forces of the vehicle suspension.

The dynamics of firing stability are usually calculated on a Digital computer and printed out in tabular form. However, when this program is used in conjunction with the simulator, the results are stored in computer memory for immediate release to the simulator. The physical arrangement of the simulator permits the occupant to fire any weapon by merely pulling the usual trigger. The command to fire is completely controlled by the man in the seat.

Sighting | Firing

FIG.10. Simulator, Firing Impulse Loop

The response to firing impulses and ride motions in the simulator were many and varied. One persistent comment was negative. People using the seat complained about the environment. They did not like the laboratory surroundings, because they were unrealistic. The "out of doors" atmosphere which is common to suspension and firing stability work was missing.

To compensate for this a visual display was

created providing a 180 degree field of view horizontally and 48 degrees in the vertical plane. A 35mm motion picture format was used to produce a "three screen" presentation. This method was selected, based upon successful tryouts of a unique projection system, developed and tailored to the simulator.

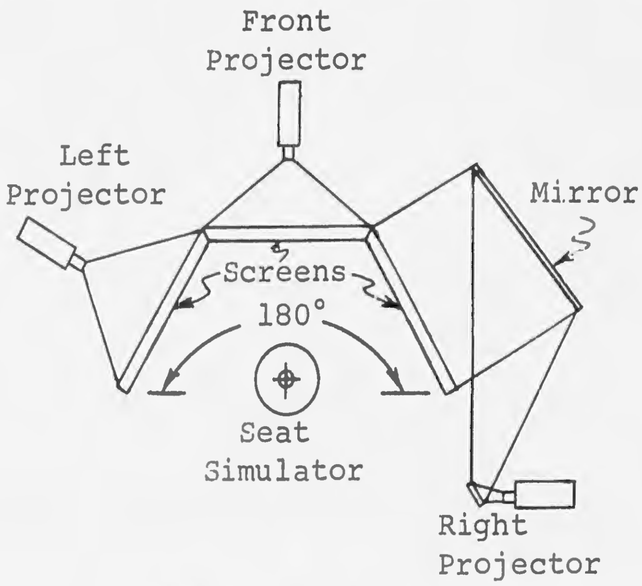


FIG.11. Visual Display System

The activity scene is photographed by three cameras and backprojected to the subject in the simulator by three synchronized-interlocked projectors. This system presents to the observer a scene that compares favorably to a view from within a combat vehicle.

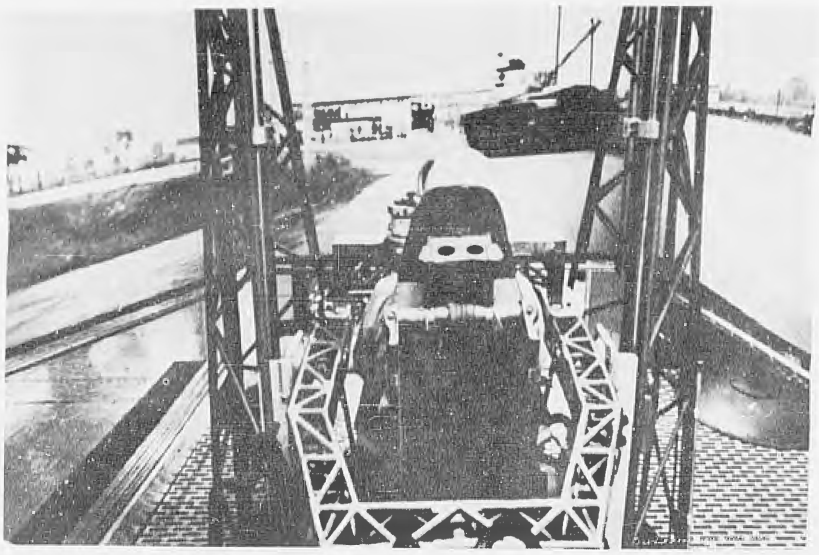


FIG.12. Visual Display

SUMMARY

The design implement created by this combined research task consists of five alloyed ingredients:

I. Equations of motion describing suspension system dynamics.

II. An original mathematical model of combat vehicle firing stability performance.

III. A versatile advanced technique for generating "real" random terrain contours.

IV. A "four degree of freedom" motion simulator machine.

V. A unique "three screen" visual display system.

The combination of these items permits investigation of suspension and firing stability characteristics without the use of physical units of the design itself. Designs may be varied easily; best combinations can be probed without expensive test rigs and test time expenditures. Vehicle design alternatives, variations and compromises, may be sorted to limit full scale work to the most promising configurations.

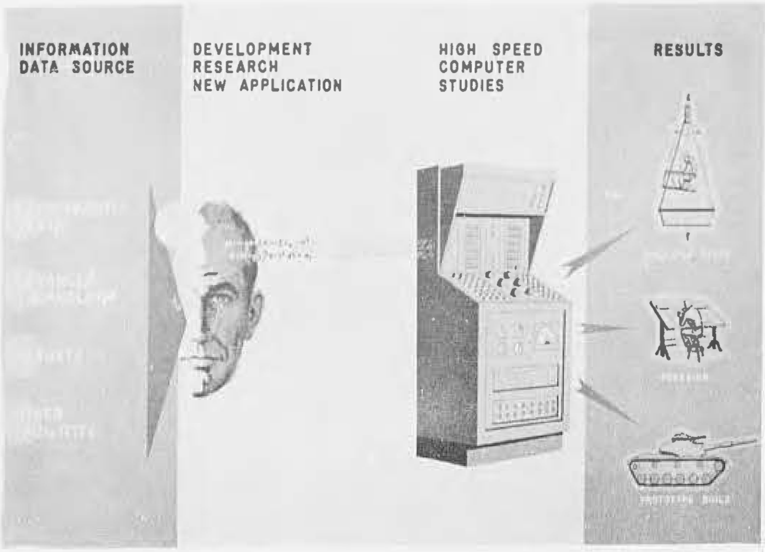


FIG.13. Integrated System

THERMAL DECOMPOSITION OF LEAD AZIDE IN VARIOUS ENVIRONMENTS

BRUNO REITZNER, JAMES E. ABEL & J.V. RICHARD KAUFMAN
EXPLOSIVES RESEARCH SECTION, PICATINNY ARSENAL
DOVER, NEW JERSEY

The thermal decomposition of metal azides including lead azide, under vacuum, proceeds at a very slow rate during the induction period. Thereafter, the reaction rate increases sharply to a point where about half of the material is decomposed, and gradually decreases again until all of the material is consumed. By plotting the fraction of the material decomposed (α) against time, sigmoid shaped decomposition curves are obtained. Reactions of this type have been studied, especially by Garner and his co-workers[1-4] who classified them as autocatalytic reactions where the product (in the case of lead azide metallic lead) is the catalyst. This assumption of an autocatalytic mechanism can be further extended to include the behavior at higher temperatures where, after a specific induction period, the reaction becomes very fast and results in a detonation.

In view of these assumptions, it seemed of interest to study the decomposition and ignition of lead azide by influencing the autocatalytic substances. Specifically, one should be able to inhibit the reaction by destroying the catalyst as soon as it is formed or enhance it by adding an external catalyst to the undecomposed material. To verify these assumptions, a program was worked out which included the following points:

a. The investigation of the effect of small amounts of water on the slow isothermal decomposition under atmospheric nitrogen and vacuum.

b. The investigation of the effect of small amounts of dry oxygen and of dry nitrogen containing iodine vapor on the slow decomposition at atmospheric pressure.

c. A comparison of the ignition behavior under vacuum and air.

d. The investigation of the effect of small amounts of metallic

silver on both the slow decomposition in vacuum and ignition in air.

The material used was colloidal lead azide with an average particle size of 7 μ. The cationic impurities were less than 0.1%. The amount of water (either adsorbed, inclosed or in a surface layer of hydrated basic lead azide) was approximately 0.2 - 0.3%.

1. The effect of small amounts of water on the slow isothermal decomposition under atmospheric nitrogen and vacuum.-The apparatus for carrying out slow decompositions under atmospheric nitrogen is shown in Fig. 1. It consisted essentially of a reaction vessel immersed in a liquid metal bath and a burette for collecting the gaseous decomposition products. The reaction vessel was filled with 50 mgs of $Pb(N_3)_2$, flushed with dry nitrogen while the outlet tube was immersed in the displacement liquid, and sealed. The water vapor pressure over the lead azide was controlled by the displacement liquids (H_2SO_4 and H_3PO_4 of various concentrations, dilute NaOH). The sealed vessel was immersed in the liquid metal bath, and the gaseous decomposition products were measured in the burette. The reaction temperature was set at 240°C because an explosion had occurred at 250°C.

Experiments under vacuum were carried out under static conditions by determining the total pressure as a function of time. The vacuum system consisted of a 4-liter bulb for collecting the gases, a Pirani gauge, a vertical sample tube (4 mm dia.) containing about 5 mgs of lead azide, and a cold trap. The system was pumped down to less than 5×10^{-5} torr and the stopcock to the pumps closed. The glass tube was then immersed in a metal bath set at 240°C. To study the effect of water, advantage was taken of the fact that lead azide contained small amounts of water which were released shortly after the lead azide was heated. By attaching a liquid nitrogen Dewar to the trap, the water could be frozen out which provided a method for comparing the decomposition curves with and without water.

The results of the experiments under atmospheric nitrogen as well as under vacuum are described elsewhere in more detail[5,6]. Only the most significant data are presented here. Figure 2 shows the decomposition curves (fractional decomposition α versus time) at 240°C. Curve 1 was obtained under dry atmospheric nitrogen using concentrated sulfuric acid as the displacement liquid. For Curve 2, a dilute sulfuric acid (spec gravity 1.513, $p_{H_2O}=3.7$ was used as the displacement liquid. Curves 3 and 4 were obtained under vacuum with and without the liquid nitrogen trap, respectively. The curves show that when the water is removed either by absorption in the concentrated sulfuric acid or by condensation in the liquid nitrogen trap, the induction times become shorter.

As a result of the removal of water, the maximum reaction rate under vacuum becomes so great as to even result in an explosion.

With water present the maximum rate is 2.08 10^{-2} min^{-1}. Under nitrogen the maximum reaction rates are 2.15 for Curve 1 and 2.5 x 10^{-2} min^{-1} for Curve 2. It has to be mentioned that water driven out of the lead azide is in contact with the lead azide and a hydrolysis equilibrium between water and hydrogen azide is established. Hydrogen azide has been found by mass spectrography of the gas sample. The mechanism of the inhibition of the reaction

$$PbN_6 \rightarrow Pb + 3N_2 \qquad (1)$$

can now be explained in the following manner: When the lead azide is heated, the water is released and the hydrolysis equilibrium is established. Both water and hydrogen azide can now react with the active lead nuclei formed by thermal excitation according to -

$$Pb + 2HN_3 \rightarrow Pb(N_3)_2 + H_2 \qquad (2)$$

$$Pb + H_2O \longrightarrow PbO + H_2 \qquad (3)$$

Both reactions result in the consumption of the metallic lead. PbO, as formed by reaction (3), is not stable with excess lead azide and is converted into basic lead azide $Pb(N_3)_2 \cdot PbO$.

Experiments made under atmospheric nitrogen using displacement liquids with a higher water vapor pressure resulted in a complete suppression of the autocatalytic reaction as shown in Fig. 3. Curve 1 was obtained under a water vapor pressure of 8.9 mm Hg, Curve 2 under 21.7 mm Hg. The displacement liquids consisted of diluted phosphoric acid of a specific gravity of 1.511 g/cc and 1.153 g/cc. There exists a proportionality between the reaction rate and the water vapor pressure. This proportionality is mainly due to a side reaction, namely the thermal decomposition of hydrogen azide formed by hydrolysis of lead azide. Hydrogen azide is decomposed into nitrogen and ammonia which reacts with excess hydrogen azide to form ammonium azide. The equation for this reaction is:

$$4\ HN_3 \rightarrow NH_4N_3 + 4N_2 \qquad (4)$$

Ammonium azide has been found as a sublimate in the colder portion of the outlet tube in weighable amounts.

The proportionality between reaction rate and water vapor pressure is no longer maintained when a sodium hydroxide solution is used as the displacement liquid because the alkaline solution absorbs a large portion of the hydrogen azide and reduces its concentration in the gas phase. The nitrogen evolution according to eq. (4) becomes, therefore, less. This is exemplified by Curve 3 which represents the data which were obtained with a sodium hydroxide solution having a water vapor pressure of 23.0 mm Hg. If the gas evolution were not due to the decomposition of hydrogen azide, Curve 3 should have a greater slope than Curve 2.

Now the question arises how much of the nitrogen evolution is due to reaction (4) and how much to the decomposition of lead azide which now is no longer catalyzed by metallic lead. By determining the amount of ammonium azide, the amount of nitrogen formed according to eq. 4 could be calculated. Subtracting this value from the final values of the gross decomposition curves and taking further into account the hydrogen evolution resulting from the destruction of the lead nuclei by water or hydrogen azide (eqs. 2 and 3) we obtain the final nitrogen values due to the non-catalytic initial reaction. Connecting these points to the origin, which is, of course, somewhat arbitrary, we obtain the dotted line 4 the slope of which corresponds to a reaction rate of 0.6×10^{-6} min^{-1}. This is approximately 1/30,000 of the maximum rate of the autocatalytic reaction.

The hydrolysis products of lead azide are two basic lead azides. These compounds can be obtained by heating lead azide in air having the usual humidity (7) or by reacting PbN_6 with PbO in evacuated vessels (8). From these experiments, the formulas of the basic lead azides have been tentatively established to be $PbN_6 \cdot PbO$ and $2\ PbN_6 \cdot 5PbO$.

2. <u>The effect of dry oxygen and of dry nitrogen containing small amounts of iodine vapor on the slow decomposition</u>.-The apparatus for performing these experiments is shown in Fig. 4. Lead azide and iodine were placed in separate compartments of the reaction tube. Oven 1 was used to heat the lead azide and oven 2 was used to heat the iodine so as to create a defined vapor pressure. The cold tube was flushed with dry nitrogen while the outlet was immersed in concentrated sulfuric acid so as to exclude water. The gases formed during decomposition were collected in a bas burette. The same arrangement was used for comparison tests under dry oxygen and dry nitrogen which were carried out without iodine. The reaction temperature in all cases was 230°; the amount of lead azide was 50 mgs; the vapor pressure of the iodine was set at 2.16 mm by heating oven 2 to 50°C. The decomposition curves are shown in Fig. 5. Curve 1 represents the data obtained under dry nitrogen. The curve resembles Curve 1 of Fig. 2 obtained at 240° under dry nitrogen. It shows an induction period of approximately 30 hours and a maximum rate of 1.65×10^{-2} min^{-1}. Curve 2 obtained under oxygen shows a marked increase in the induction period (410 hrs). The maximum reaction rate was substantially the same as under dry nitrogen. The long induction period shows that the autocatalytic reaction is inhibited by the oxygen. Gradually the oxygen in the tube will be consumed by reaction with the lead nuclei and finally the autocatalytic reaction will no longer be inhibited. The reaction rate during the induction period was found to be 1.1×10^{-6} min^{-1} which is smaller by a factor of 10^4 than the maximum rate under dry nitrogen. This is in agreement with the results described in the first chapter. When lead azide is decomposed under nitrogen containing iodine vapor (Curve 3), the autocatalytic reaction is likewise suppressed. Instead of a sigmoid

type decomposition curve, a straight line is observed. The rate of the reaction is 4.54×10^{-4} min^{-1} which is smaller by a factor of 30 than the maximum rate under dry nitrogen. It shows that the iodine likewise destroys the lead nuclei. However, if this would be the only effect, we should expect a rate similar to that found under oxygen during the induction period. We, therefore, have to assume that the iodine not only reacts with the lead nuclei formed by thermal excitation, but also with the azide group. The linearity of Curve 3 which suggests a zero order reaction can be explained by the assumption that the rate determining step is the diffusion of the iodine from the reservoir to the lead azide. The final solid decomposition product in the last case was lead iodide.

3. A comparison of the ignition behavior of lead azide under vacuum and air.-The ignition times τ as a function of the temperature were determined with an ignition apparatus which consisted of a hot plate enclosed in a vacuum bell jar. A multi-shot dropping mechanism was used to drop about 50 samples of lead azide (average weight 1.1 ± 0.1 mg) on the hot plate without breaking the vacuum. When the sample was dropped on the hot plate a light beam focussed across the hot plate to a photomultiplier was interrupted which triggered a timing mechanism. The light flash emitted by the explosion caused a change in the voltage output of the photomultiplier tube which was used to stop the timer. The same apparatus could be used for experiments in air. The apparatus will be described in more detail in a forthcoming publication.

Work with this apparatus has been started only recently, and the data are not yet complete. The data already obtained show that the ignition curves in air are cut off at a temperature of about 380°C. Below this temperature the material does not ignite any more and is converted into the second basic lead azide which has a minimum ignition temperature of more than 500°C. In vacuum, however, the curve could be extended towards lower temperatures. So far, we have made measurements down to 355°C, but did not yet attempt to go lower. According to Curve 3 of Fig. 2 which was obtained with 5 mg. of lead azide it is apparent that we can go down as far as 240°C. The explanation for this is the same as given before. Below a certain temperature, the rate of lead nucleus formation is too low as to overcome the rate by which they are oxidized, and the autocatalytic reaction which leads to explosion cannot develop. The minimum ignition temperature of 380°C in air does not agree with the results of other workers who found a value of about 350°C. The reason for this discrepancy lies in the fact that we used samples containing only 1 mg. of lead azide as compared to 10 to 20 mg. used by previous authors. We, therefore, have a higher bulk surface-mass ratio which facilitates the oxidation of the lead nuclei even if they are formed at a higher rate. An investigation of the mass effect is under way.

4. <u>The effect of small amounts of silver on the slow decomposition of lead azide in vacuum and on the ignition in air.</u>-For the following experiments the vacuum system and the ignition apparatus described earlier were used. Silver was deposited on the lead azide by reducing a methanol-water solution of silver nitrate in which lead azide was suspended by means of hydrazine hydrate.

The early stages of the decomposition curves in vacuum at 240°C are shown in Fig. 6. The curves which represent the initial stage of the decomposition were obtained by condensing the outcoming water-hydrogen azide mixture by means of a liquid nitrogen trap. Curves 1 and 2 show the decomposition of lead azide coated with 1.0 and 10.0 atom % of silver, respectively. Curve 3 shows the behavior of untreated lead azide, and curve 4 of lead azide treated with hydrazine hydrate alone. It can be seen that the induction period is reduced by the addition of silver as compared with that of the untreated lead azide. The treatment with hydrazine hydrate alone does not produce any accelerating effect. These experiments were supplemented by ignition tests in air. Figure 7 shows the ignition times τ at various temperatures plotted against the percentage of silver. The ignition times were decreased with small amounts of silver (minimum at about 1.0 at -%). At higher concentration, the silver acts as a heat sink causing the ignition time to rise again.

Conclusions

It can be concluded from the present experiments that reagents capable of reacting with lead such as water, hydrogen azide, oxygen and iodine increase the induction period of the autocatalytic decomposition of lead azide. On the other hand, the presence of a metal (it does not have to be lead) enhances the decomposition and ignition by its catalytic action. It is planned to use other metal azides for studying these environmental effects so as to be able to generalize the conclusion that oxydizing agents which are capable of reacting with the metal nuclei formed by thermal excitation are likewise capable of suppressing the reaction. Even with the present results we feel that we can contribute some facts to the theory of auto-accelerated reactions. Two views are currently discussed: Semenoff (9) assumes that the intermediate products formed during the first stage of the reaction activate neighboring molecules at the cost of energy liberated during the decomposition thus promoting the development of a chain. These intermediate products and not the final products act as catalysts. Garner[3], however, assumes that the final product, in the case of metal azides the corresponding metal, is the catalyst. The results of the present experiments can be much easier explained by Garner's theory. Suppose the active species is some sort of a germ nucleus (this term may include F-centers or colloidal metal) then its reaction with oxygen or iodine should release additional energy with the result that the adjacent molecules should be activated to a larger extent and the reaction should be accelerated. This,

however, is not the case. Likewise, the catalytic effect of an externally deposited metal which is not present in an active form can be better explained according to Garner's theory.

References

1. W.E. Garner and A.S. Gomm, J. Chem. Soc., 2123 (1931)

2. W.E. Garner, A.S. Gomm and H. R. Hailes, ibid., 1393 (1933)

3. W.E. Garner, "Chemistry of the Solid State", London 1955, Ch. 7 and 9.

4. W.E. Garner, Proc. Roy. Soc. (London), A246, 203 (1958).

5. B. Reitzner, J.V.R. Kaufman and E.F. Bartell, J. Phys. Chem. in print.

6. B. Reitzner, ibid. 65, 948 (1961).

7. M. Stammler and J.E. Abel, Advances in X-Ray Analysis, Vol. 4, 130 (1960) and Nature 192, 626 (1961)

8. B. Reitzner and J.E. Abel, Advances in X-Ray Anal, Vol. V (1961) in print.

9. N. Semenoff, Chemical Kinetics and Chain Reactions, Clarendon Press pp 437 f (1935).

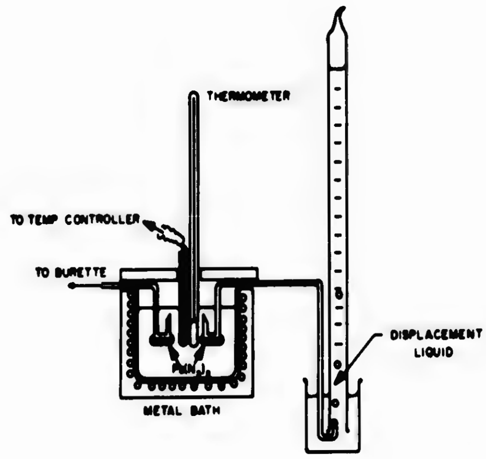


FIG.1 SLOW DECOMPOSITION APPARATUS

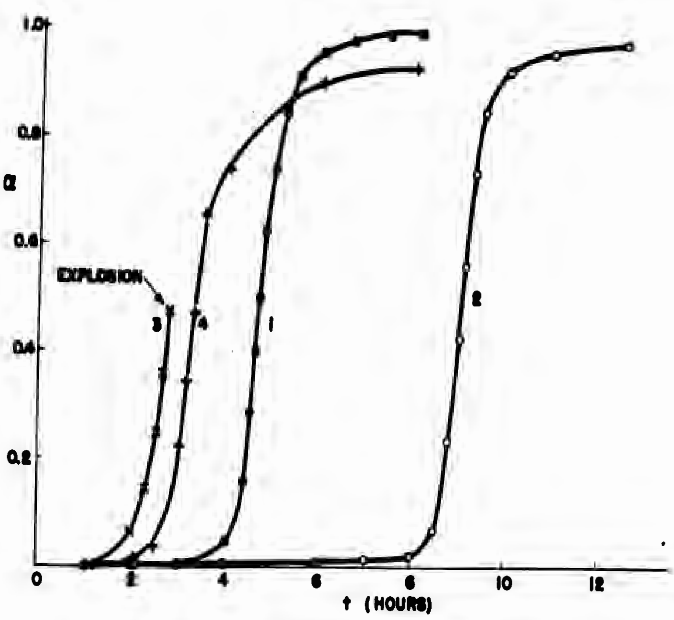


FIG.2 DECOMPOSITION CURVES OF $Pb(N_3)_2$ UNDER NITROGEN (1-2) AND VACUUM (3-4)

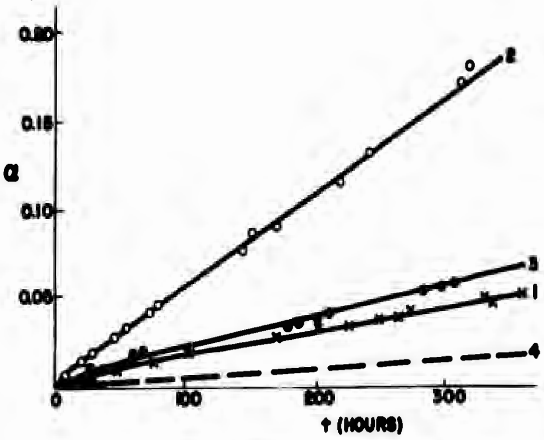


FIG. 3 DECOMPOSITION CURVES OF $Pb(N_3)_2$ UNDER HUMID NITROGEN

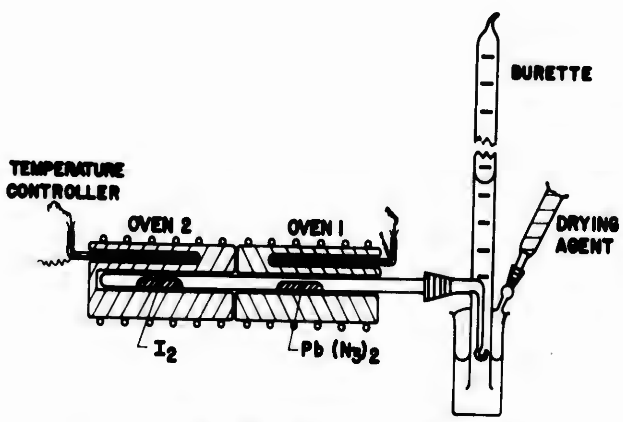


FIG. 4 SLOW DECOMPOSITION APPARATUS FOR IODINE EXPERIMENTS

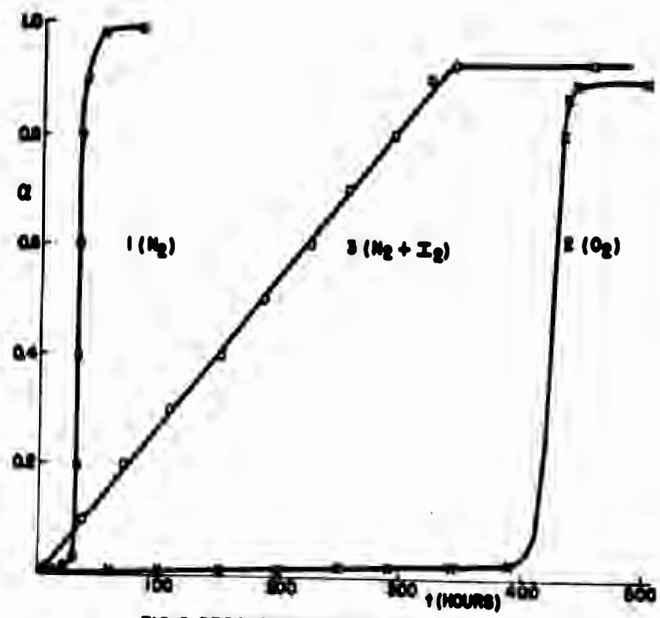


FIG. 5 DECOMPOSITION CURVES OF $Pb(N_3)_2$ UNDER N_2, O_2, AND $N_2 + I_2$

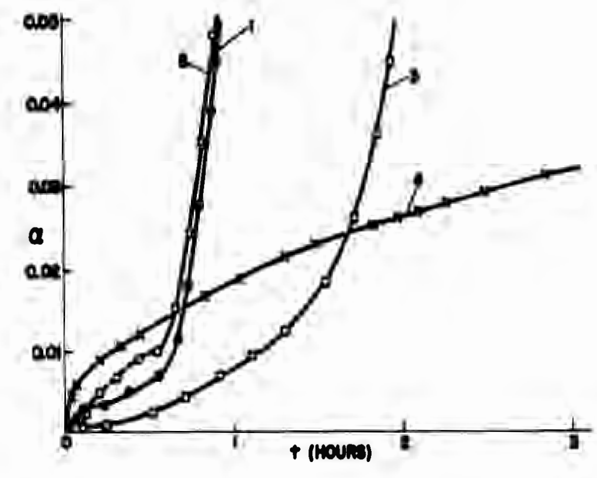


FIG. 6 DECOMPOSITION CURVES OF UNTREATED SILVER COATED $Pb(N_3)_2$ (INITIAL STAGE)

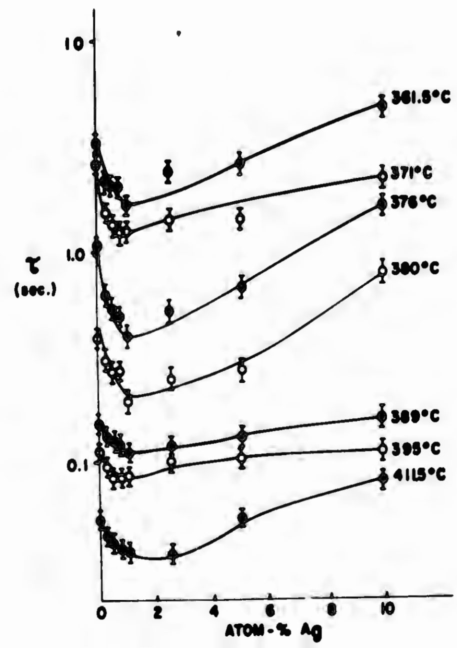


FIG. 7 IGNITION TIMES OF $Pb(N_3)_2$ COATED WITH SILVER

THE SIGNIFICANCE OF SPECTROSCOPICALLY DETERMINED PLASMA JET TEMPERATURES

T. G. ROBERTS
U.S. ARMY ORDNANCE MISSILE COMMAND
REDSTONE ARSENAL, ALABAMA

The measurement of plasma temperatures is of considerable practical importance and much effort has been expended by many investigators to develop and improve techniques for the determination of plasma temperatures. Several excellent review papers (1-3) have appeared in the recent literature and that of Margenau and Lewis (4) is particularly noteworthy. In spite of (or perhaps because of) this effort, there exists considerable confusion regarding these temperatures. It is important to recognize that the word "temperature" may have many different meanings which depend on the nature of the system under study and the methods used to determine these temperatures.

If a source is in thermal equilibrium, then the problem of temperature determination is straightforward, and any of the equations used to define temperature can be used to calculate the value of the temperature. Each calculation in this case would lead to the same number for the temperature no matter which method is used. However, thermal equilibrium conditions are rarely found in plasma generated in laboratories and the concept of temperature in a plasma is complicated by departures from thermal equilibrium (9). It should be remembered that in the derivation of these equations the Maxwell-Boltzmann statistics are used and it is assumed that the coupling between the different degrees of freedom is such that the equipartition of energy holds.

An isolated system which emits continuous radiation or spectral lines can not be truely in thermal equilibrium since the principle of detail balancing does not hold because of the escaping radiation. It often happens that the departures from equilibrium are such that the various degrees of freedom have different temperatures, i.e. the equipartition of energy does not hold, but the

energy within one or more degrees of freedom is distributed in accordance with Maxwell-Boltzmann statistics. This result is typical of most plasma jets. In this type of system the temperature has no real meaning and the fundamental question of energy distribution must be answered directly. A temperature number is something that is useful only within the limits of its definition. If a temperature can be defined so as to give the velocity distribution of some constituent of the plasma the populations of the rotational energy levels, the vibrational energy levels, or the atomic energy levels, or to give the power radiated per unit frequency, the power radiated in the form of bremsstrahlung, or the total power radiated per unit surface area, then it is a useful parameter. But, if temperature does none of these things then we must forget temperature and try to describe the system in terms of density, pressure and enthalpy or try to find a distribution function which can be used.

This complex nature of nonequilibrium systems has lead to many different "temperatures" which are in use today. It is the purpose of this report to explain the significance of these temperatures in a semi-qualitative manner by giving a brief review of the formulas which may be used to define temperature in the case of thermal equilibrium. The manner in which these formulas are used to define the various temperatures of a plasma which is not sufficiently in equilibrium for the different degrees of freedom to be coupled will also be shown. Examples of three of the techniques will be given utilizing measurements from the AOMC carbon electrode plasma jet.

Two limiting cases will be considered for systems which are in equilibrium and for systems which are not in equilibrium. These cases are: 1) the plasma is "optically thick", spatially homogeneous, and in a steady state, and 2) the plasma is spatially homogeneous and in a steady state, but is "optically thin". It is only necessary to consider optical radiation to clarify the meaning of the different temperatures used in plasma physics. Two basic methods which may be utilized are to make measurements on the radiation emitted by and absorbed by plasmas. The problem can be further simplified by considering only the case of emitted radiation. Measurements on the emitted radiation from plasmas can be used to identify the atomic species present in the plasma, to estimate the temperatures of the plasma, and under certain conditions to estimate the ion or electron density. These measurements, of course, possess the unique advantage of permitting estimates of these parameters without disturbing the plasma under study. It is not generally possible however, to make a point by point temperature exploration of a spatially inhomogeneous plasma by the measurement of the emitted radiation.

EQUILIBRIUM PLASMAS - A. Optically Thick: When the plasma is spatially homogeneous in a steady state, and is "optically thick", the emitted radiation is given by Planck's Law

$$U(\nu)d\nu = \varepsilon 2h2^{3} [\ C^{2}(\exp h\nu/kT-1)\ ]^{-1} d\nu \tag{1}$$

where $U(\nu)$ is the radiation density in ergs/cm^2, ν the frequency, ε the emissivity, and the other symbols have their usual meaning. For a black body, $\varepsilon=1$; for a gray body, ε=a constant <1; but for most substances the emissivity is a function of both ν and T; i.e., $\varepsilon=\varepsilon(\nu, T)$.

Integration of equation (1 over all frequencies for a black body given Stefan-Boltzmans Law: $E = \sigma T^4$

$$E = \sigma T^{4} \tag{2}$$

where E is the total energy radiated per unit surface area of the source and $\sigma = 5.67 \times 10^{5}$ erg cm^{-2}sec^{-1} deg^{-4}.

For low frequencies ($h\nu/kT<<1$), the Rayliegh-Jeans approximation to Planck's Law is obtained:

$$U(\nu)d\nu = 2\varepsilon\nu^{2}kTd\nu/C^{2} \tag{3}$$

For high frequencies [$\exp(h\nu/kT)>>1$] Wein's approximation to Planck's Law applies:

$$U(\nu)d\nu = 2\varepsilon h\nu^{3}(C^{2}\exp h\nu/kT)^{-1} d\nu \tag{4}$$

Wein's displacement law results from the frequency of maximum radiation:

$$T = h\nu_{max}/2.82k \tag{5}$$

B. Optically Thin: For the "optically thin" source information can be gained from the emitted radiation by use of the following formulas:

The intensity of a spectral line is given by:

$$I_{m'm''} = (N/B)\ g_{m'}A_{m'm''}\exp(-E_{m'}/kT) \tag{6}$$

with

$$B = \sum_{i} g_i \exp(-E_i/kT) \tag{7}$$

where N is the number of atoms per cm^3, $g_{m'}$ the statistical weight of the upper state, $E_{m'}$ the excitation energy of the upper state, $A_{m'm''}$ the transition probability, and B is the partition function.

It is not necessary to measure the absolute intensity of a given spectral line as the ratio of the intensities of two lines leads to:

$$I_{m'm''}/I_{n'n''} = C \exp - (E_{m'}-E_{n'})kT \tag{8}$$

with

$$C = g_{m'}A_{m'm''}/g_{n'}A_{n'n''} \tag{9}$$

A plot of $\ln(I_{m'm''}/CI_{n'n''})$ versus the difference of the energies yields a straight line with slop equal to $-1/kT$. Similar expressions can be given for the rotational and vibrational structure of the band spectra of molecules.

The degree of ionization is related to the temperature through Saha's equation:

$$\frac{N_i N_e}{N_o} = \frac{2g_i}{g_o} \frac{\pi^{3/2}(2m_e kT)^{3/2}}{h^3} \exp \frac{-X}{kT} \tag{10}$$

where N_i is the ion density, N_e the electron density, N_o the neutral particle density, and S is the effective ionization potential which is determined from the principle quantum number for which the partition function is truncated. From equations (6 and (10, the ratio of the intensities of a line from the atom to a line from its ion is obtained.

$$\frac{\overset{+}{I}_{n'n''}}{I_{m'm''}} = \frac{\overset{+}{\nu}_{n'n''}\overset{+}{A}_{n'n''}\overset{+}{g}_{n'}}{\nu_{m'm''}A_{m'm''}g_{m'}} \frac{2(2\pi m_e kT)^{3/2}}{N_e h^3} \exp - \frac{X}{kT} \tag{11}$$

where ν, A, g, and X are in principle obtained from spectroscopic tabulations and if N_e is known then T can be evaluated.

The temperature can also be evaluated from the rate of energy emission per unit wavelength in the interval λ to $\lambda + d\lambda$ due to bremstrahlung as follows:

$$dP_\lambda = C^1 g N_e \Sigma(N_i Z^2) T^{-\frac{1}{2}} \lambda^2 \exp(-hC/\lambda kT) d\lambda \tag{12}$$

where g is the gaunt factor which for high temperatures approaches the value $2 \times 3^{\frac{1}{2}}/\pi$ and is generally taken to be unity. If λ is expressed in angstrom units and T is kev then for hydrogen the constant C^1 equals 6.01×10^{-30}.

Under certain conditions the temperature can be estimated from the Doppler broadening where the width of the line at the point of half maximum intensity is given by:

$$\Delta\lambda_{\frac{1}{2}} = 1.67 (\lambda/C)(2kT/m)^{\frac{1}{2}} \tag{13}$$

where m is the molecular weight, and the constant is given for d.g.s. units with T in ^{o}K. In order to use equation (13 the broadening must be due to the Doppler effect, as will be the case if the line profile is gaussian.

For every spectral line of every element there is a temperature at which the intensity of the line reaches a maximum. At higher temperatures the effect of ionization causes the intensity to drop off, and at lower temperatures the intensity is proportional to the temperature. The intensity of a spectral line is given by equation (6 as a function of the temperature and the particle density. Equation (10 gives the particle density as a function of the temperature. These two equations may be solved for the intensity as a function of the temperature from which the temperature of maximum intensity may be calculated.

NONEQUILIBRIUM PLASMAS-A. Optically Thick: When the plasma is spatially homogeneous, in a steady state, and is "optically thick" equations (1 through (5 are used to define a color temperature, a brightness temperature, and an effective temperature. The color temperature is defined as the temperature at which a blackbody emits radiation having the same ratio of radiant intensity at λ_1 and λ_2 as the source under study. The brightness temperature is defined as the temperature at which a blackbody would emit radiation in the chosen spectral range (which must be stated) of the same intensity as that emitted by the plasma under study. The effective temperature is defined as the temperature of a blackbody whose surface radiates exactly the same amount of energy per unit area as the plasma under study.

B. Optically Thin: The "optically thin" case is by far the most interesting because it is the most often encountered, and the most complicated. Here, equations (6 through (13 are used to define excitation, rotational, vibrational, electron, ion, and kinetic temperatures.

The excitation temperature (1, 3, 5, 8, 10-15) which gives the populations of the atomic energy levels is defined by equations (6 or (8. It is preferable to use equation (8 and a large number of lines to insure that the energy levels are populated according to some exponential partition function like the Maxwell-Boltzmann partition function.

Rotational and vibrational temperatures (5, 8, 16-22) which give the populations of the rotational and vibrational energy levels are defined by equations similar to equations (6 and (8.

The ionization temperature (3, 5, 8, 23, 24) is defined by equations (10 or (11 and should lead to the degree of ionization. Ordinarily equation (11 is used and the ratio of the intensities of more than two pairs of lines from the atom and its ion should be

checked to insure that the ionization temperature measured is in fact proportional to the degree of ionization. For low temperature diatomic gases it may be convenient to define a disassociation temperature which can be done using a formula similar to the Saha equation (25).

The electron temperature (3, 5, 8) which is proportional to the average kinetic energy of the electrons is defined by equation (12. Again it is preferable to take the ratio of the power radiated per unit wavelength in the form of bremstrahlung for two different regions of the spectrum such as the visual and the soft X-ray region.

Equation (13 is used to define the kinetic temperature (8, 26) of the atoms or ions which emit the line utilized. This method is the only one available for directly measuring the kinetic temperature.

All other temperatures (5, 8, 27-30) which may be encountered in the literature such as the N_a line reversal temperature etc., comprise one or more of the above temperatures named so as to be descriptive of the method of measurement.

The temperature which is determined from the maximum intensity of a spectral line (5, 31-33) is the same as the excitation and ionization temperatures if these two are equal. Otherwise, this temperature is some kind of an average of the two.

It was mentioned that all methods of computing temperature give the same results for systems in thermal equilibrium. However, many systems are not in thermal equilibrium and the different temperatures which one may measure depend upon the physical phenomenon which one wishes to describe. Whenever one of the above definitions is used to determine a temperature for a thin plasma, such as an excitation temperature, it is implicitly assumed that the excitation energy is distributed among the excited states in accordance with Maxwell-Boltzmann statistics. Experience has shown that this assumption must be verified experimentally, for abnormal excitation mechanisms lead to erroneous results. Also, in the excitation temperature measurement the sources which are generally encountered are neither optically thick nor optically thin. In general they are thin in the continuum where no lines occur and thick to the resonant radiation. Therefore, it is necessary to measure the self-absorption or to correct for it in some manner, e.g. using the theory of Cowan and Deike (34), before the numbers obtained will be indicative of the conditions within the plasma under study.

ILLUSTRATIVE EXAMPLES: Three of the methods which have been used in this laboratory to measure temperatures in a plasma jet have been selected for the purpose of illustration. These methods are: (1) the two line method for determining the excitation temperature (35), (2) the use of continuous radiation from an object in the jet stream to determine a brightness temperature (36), and

(3) the use of the maximum intensity of a spectral line for the determination of an excitation-ionization temperature (32, 33).

A. The Two Line Method: An iron impurity was added to an argon plasma jet. It was assumed that the excitation temperature of the iron is the same as that of the argon so that the transition probabilities of the iron lines could be computed from the formula (37):

$$A_{n'n''} = \frac{g_{n''}}{g_{n'}} \quad \frac{8\pi^2 e^2}{me^2} \quad f_{n''n'}$$

where e is the electronic charge, c the velocity of light, and $f_{n''n'}$ is the ordinary absorption oscillator strength. The two Fe lines used were $\lambda = 3902.95$Å and $\lambda = 4005.25$Å. The relative f values for these lines have been determined experimentally (38).

Figure 1 shows this region of the spectrum, and measurement of the intensities from this plate gives, through the use of equation (8, an excitation temperature of $T_{\lambda\lambda}3902.95\text{-}4005.25 \approx 8000°K$. This temperature was measured at the center of the jet and at the nozzle exit when the power delivered to the jet was 13 KW.

For this measurement, lines were chosen which were close together on the spectrum, and of comparable intensities. They were also chosen because they were weak lines which tends to minimize the effect of self-absorption. However, no attempt was made to determine whether or not the excitation temperature of the iron was in equilibrium with the excitation temperature of the argon.

B. The Brightness Temperature of an Ablating Model: The system considered here consists of a solid immersed in a plasma stream flowing slightly below Mach nr. 1. The character of the stream is such that it will transfer about 500 watts/cm^2 to a cold probe at the model test position. When ablation begins, the model has a high surface temperature, and a sharp temperature gradient exists in the model. As ablation proceeds the surface moves as material is removed from the model. Simultaneously, there is a rather dense, high temperature gas envelope surrounding the model. This gas is a mixture of varying composition made up of the plasma jet gas and the evaporated materials from the model.

The brightness temperature is measured spectroscopically by comparing the radiation from the solid model to that from a standard tungsten lamp. The experiment was arranged so that the radiation from the lamp traveled the same optical path as that from the model. Since the emissivity for tungsten as a function of T and λ is known (39) for the temperature and spectral range used, the radiation from the lamp may be corrected to that from a blackbody.

The experimental arrangement is shown in Figure 2, Figure 3 shows a typical plate on which the continuum of the

spectrum of the model (in this case a 3/4 inch graphite hemisphere), the lamp spectrum, and the argon plasma jet spectrum are shown. Figure 4 shows densitometer traces of some of the spectra of Figure 3 where the numbers on the traces correspond to the numbered spectra of Figure 3. The densitometer traces show a peak background radiation near the 8000 angstrom position of the curve. The general background radiation curve is evident from the standard lamp tracing and the results from a composite of the lamp radiation, the photographic plate sensitivity, and the various attenuations in the optical system. Since, there was no strong band structure near this peak intensity on the plate, the 8000 angstrom unit region was selected to calibrate and measure the temperature of the ablating surface.

The standard lamp was operated at various temperatures, and by relating these temperatures to the densitometer traces, a temperature could be estimated for which the plate transmission was the same for both the standard lamp and the model. Correction of this particular lamp temperature to a blackbody temperature gives the brightness temperature for the model. This procedure leads to

$$T_{\lambda = 8000 Å} = 2140^{o}K$$

for the brightness temperature of the graphite model.

C. The Temperature of Maximum Intensity: The simultaneous solution of Saha's equation (10, the equation of charge equilibrium, and Dalton's law of partial pressures for an argon plasma which behaves as an ideal gas leads to the following equation:

$$dN_o/dT = N_o E_{n'} /kT^2 \qquad (14$$

which defines the condition to be satisfied at the intensity maximum of a spectral line.

Figure 5 shows a graphical solution of this equation for the 4289.09Å argon line. The point of intersection of these two curves gives the values of dN_o/dT and $N_o E_{n'} /kT^2$ which are equal and identifies the temperature of maximum spectral intensity.

In this experiment, the image of the argon plasma was focused upon the entrance slit of a 3.4 meter Ebert spectrograph which has essentially no astigmatism in its optical system. This arrangement results in an accurate correspondence between points in the plasma and the spectral lines. Figure 6 shows a spectrum of an argon plasma jet in air, the image corresponding to a section of the jet 5 millimeters downstream from the nozzle. The 4289.09Å argon line, which is used to illustrate the data, is indicated by an arrow in the figure. The apparent location of the radiation

maximum is easily distinguishable.

Since Saha's equation was used in the above calculations, the temperature of maximum intensity is different for different pressures. This effect is illustrated in Figure 7 where the temperature of maximum intensity for the first three lines of the hydrogen Balmer series is given as a function of the pressure. The temperature gradient in the plasma jet had to be such that the intensity passes through a maximum, because this is the only calibration point for this method. Other temperatures along the spectral line and therefore across the plasma jet must be determined relative to this one.

It should be noted that the spectrum of Figure 6 is a one dimensional projection of a two dimentional cross section of the plasma jet. This fact however, does not influence the value of the temperature found for the maximum emission of a spectral line but does have an effect on the radial position of the determined temperature. In order to correct the radial position it is necessary to determine the radial distribution, I(r) of emission from the observed distribution, I(x). This problem has received extensive treatment and references 5, 8, and 40 may be consulted for details.

CONCLUSIONS: In certain cases the electron temperature may be less than the ionization temperature, while the ionization temperature may be greater than or less than the excitation temperature. In general the effective, color, brightness, atomic excitation, ionization, ion, electron, and etc., temperatures are different for a plasma. Each of these temperatures describes a different phenomenon and each contributes to our information about the physical conditions of the plasma.

An effort has been made to clarify the meanings of the different temperatures used in plasma physics. To accomplish this clarification use has been made of the laws which govern the emitted radiation from sources in equilibrium. These laws have been extrapolated to systems which are not in equilibrium by assuming that the energy contained in any one degree of freedom is distributed within this degree of freedom in accordance with Maxwell-Boltzmann statistics.

Examples of the measurement of three of these temperatures have been included as illustrations. It has been pointed out that only the principles involved in the determination of these temperatures are stressed in this report. Before any of these measurements are made, the referenced literature should be consulted for details and expected difficulties.

REFERENCES

1. Petrie, W., AM. J. Physics, 16, 1948.
2. Pearce, W. J., WADC-TR-59-364, February 1960.
3. Lochte-Holtgreven, W., Schall, R., and Wecken, F., Rep. Prog. Phys., 21, 1958.
4. Margenau, A., and Lewis, M., Revs. Mod. Phys., 31, 569 (1959).
5. Temperature, Its Measurement and Control in Science and Industry, Vol. II, (American Institute of Physics, Reinhold Publishing Corp., New York, 1955).
6. Lochte-Holtgreven, W., Appl. Sei. Res., Section B, Vol. 5 1955.
7. Temperature, Its Measurement and Control in Science and Industry, (American Institute of Physics, Reinhold Publishing Corp., New York, 1941).
8. Optical Spectrometric Measurements of High Temperatures Edited by P. J. Dickerman, The University of Chicago Press 1961.
9. Krysmanski, K. H., Ann. Physik, 7, 6, 1960.
10. Elenbaas, W., "The High Pressure Mercury Vapor Discharge", Interscience Publishers, New York, 1951.
11. Hefferlin, R., J. Opt. Soe. of Am., 49, July 1959.
12. Hefferlin, R., Cobb, B., Hall, D., and Lehman, C., Ap. J., 132, July 1960.
13. Howard, R., Ap. J., 127, 1958.
14. Belousova, I. M., and Gurevish, D. B., Opt. & Spec. X, March 1961.
15. Pearce, W. J., Conference on Extremely High Temperature, pp 123-134 (John Wiley and Sons, Edited by Fischer and Mansur, 1958).
16. Dieke, G. H., Data on Equilibrium Conditions-Rotational Temperatures, (Physical Measurements in Gas Dynamics and Combustion, Edited by Ladenberg, et. al. Princeton University Press, 1954).
17. Duffendack, O. S., Revans, R. W., and Roy, A. J., Phys. Rev., 45, June 1934.
18. Diniak, A. W., Roth, C., Korman, S., and Sheer, C., Vitro Labs. Rpt. KLX-10113, August 1958.
19. Gaydon, A. G., and Wolflard, H. G., RoyalSoe (London), Proc., Vol-194 169, 1948.
20. Gaydon, A. G., and Wolfhard, H. G., Royal Soe. (London), Proc., Vol-199, 89, 1949.
21. Gaydon, A. C., and Wolfhard, H. G., Royal Soc. (London), Proc., Vol-201, 561, 1950.

22. Herzberg, G., Spectra of Diatomic Molecules, D. Van Nostrand Company, 1950.
23. Menzel, D. H., Nat. Acdy, Sci., Proc., 19, 1933.
24. Dewan, E. M., Phy. of Fluids, 4, June 1961.
25. King, L. A., Appl. Sci, Res., Section B, Vol. 5, 1955.
26. Burger and VonCittert, Zeit. Physik, 51, 638, 1928.
27. Penner, S. S., Am. J. Phy., 17, September 1949.
28. Penner, S. S., Am. J. Phy,, 17, November 1949.
29. Bitter, F. and Waymouth, J. F., J. Opt. Soe. of Am., 46 1956.
30. Strong, H. M., Bundy, F. D., and Larson, D. A., "Temperature Measurement on Complex Flames by Sodium Line Reversal and Sodium Line Intensity Contour Studies", Third Symposium on Conbustion Flame and Explosion Phenomena, The Williams & Wilkins Co., Baltimore, Maryland, 1949.
31. Going, W., Zeit. Far Physik, Vol. 131, 1952.
32. Quarterly Research Review No. 17, Ordnance Missile Laboratory Division, ARGMA, Report 2A19, 1 Dec 58.
33. Quarterly Research Review No. 23, Ordnance Missile Laboratory Division, ARGMA Report TN 1C1N-23, 1 Dec 59.
34. Cowan, R. D., and Cieke, G. H., Rev. Mod. Phys., 20, 418-455, April 1948.
35. Quarterly Research Review No. 17, Ordnance Missile Laboratory Division, ARGMA Report 2A17, 1 June 1958.
36. Quarterly Research Review No. 21, Ordnance Missile Laboratory Division, ARGMA Report 1C21N-1, 1 June 1959.
37. Aller, L. H., "The Atmosphere of the Sun and Stars", The Roland Press Co., New York, 1953.
38. King, R. B., and King, A. S., Atrophys, J.,
39. Lewis, B. and Von Elbe, G., Combustion Flames and Explosions of Gases, pp. 683-684, Academic Press Inc., New York, 1951.
40. Nagler, R. G., JPL-TR-32-66, January 1961.

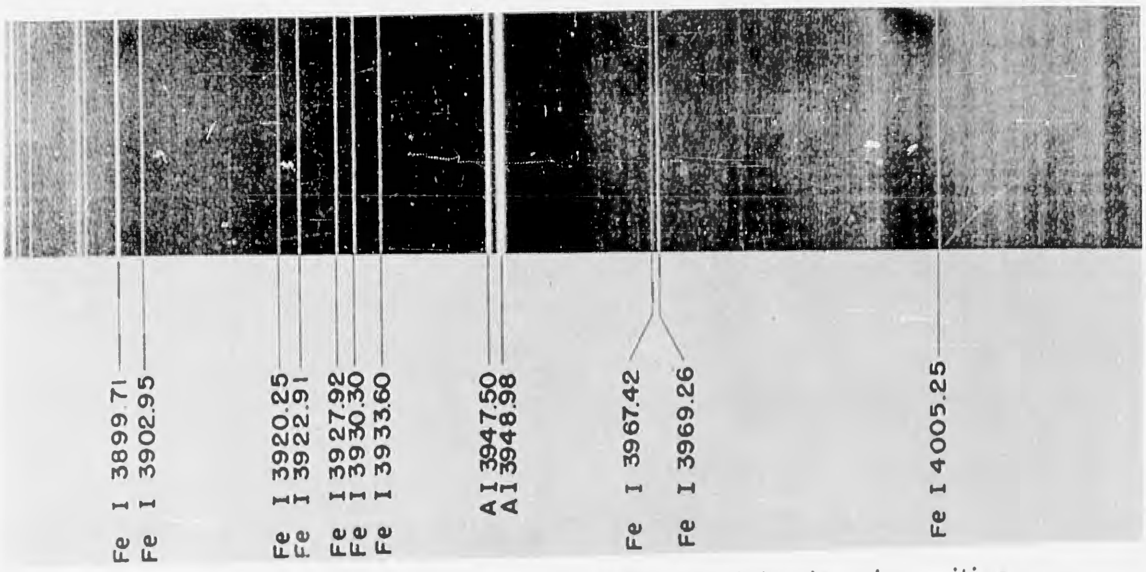


FIGURE 1 - Spectra of an argon plasma containing iron impurities

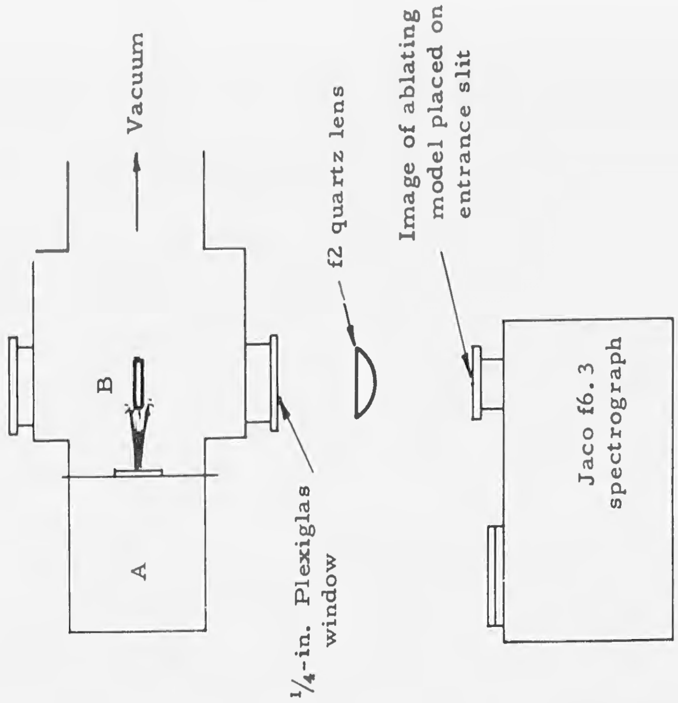


FIGURE 2 - Brightness Temperature Measurement System

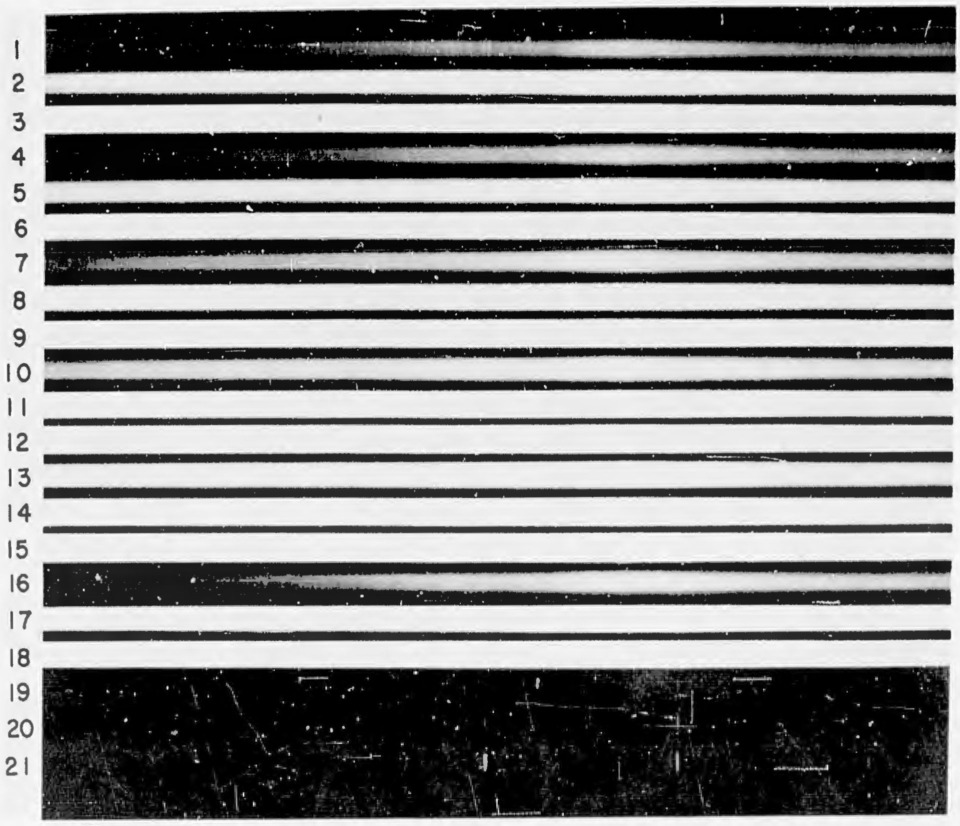


FIGURE 3 - Typical spectrograph for measurement of brightness temperature

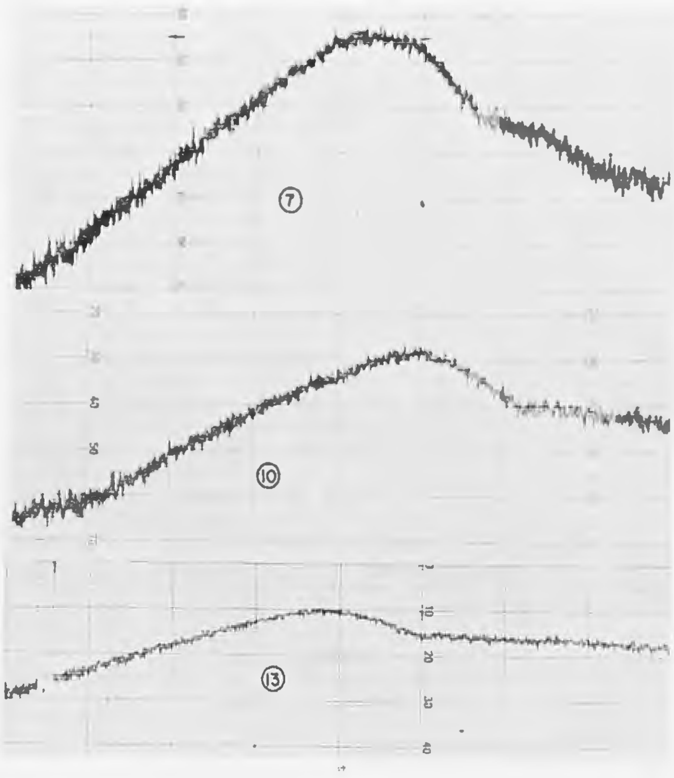


FIGURE 4 - Demsitometer traces of certain spectra of Fig. 3 numbers in O correspond to the numbers on the spectra in Fig 3

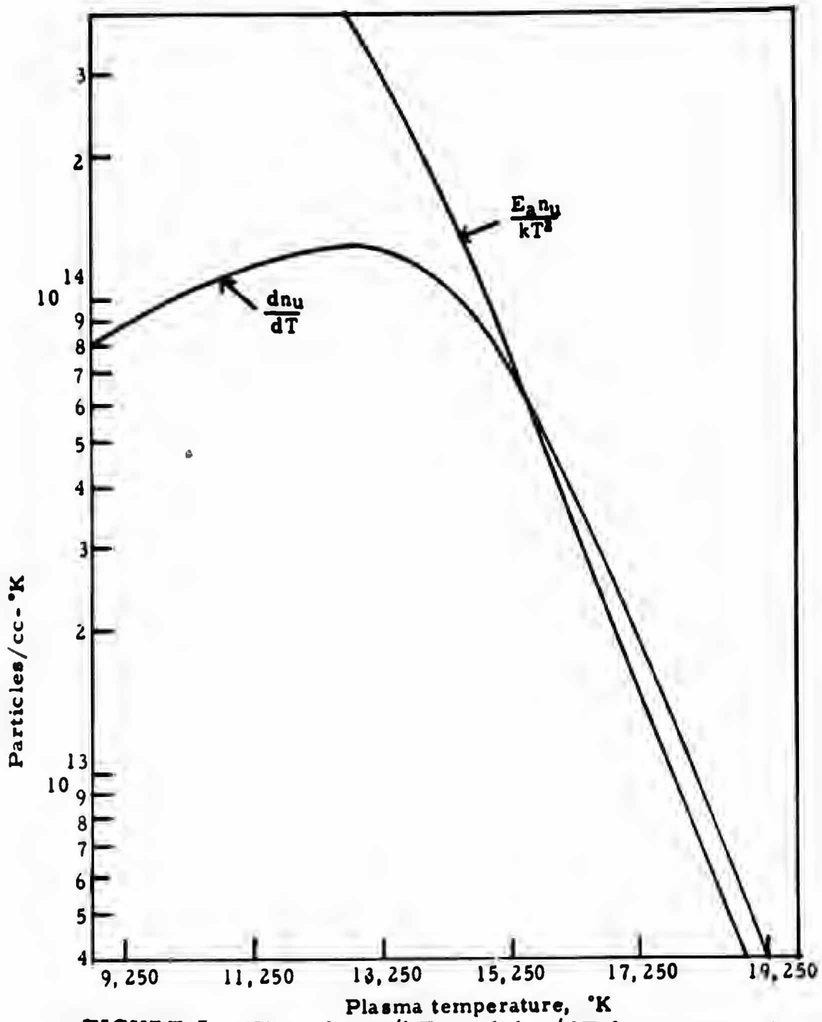


FIGURE 5 - Plot of En_o/kT^2 and dn_o/dT for argon at 1 atmosphere pressure

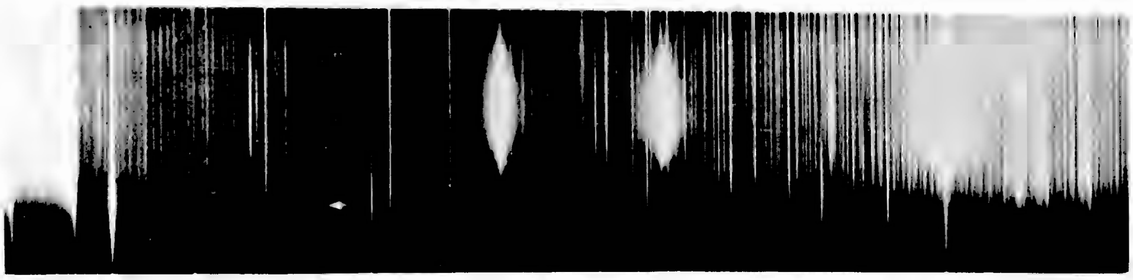


FIGURE 6 - Spectrum of argon plasma operating in air

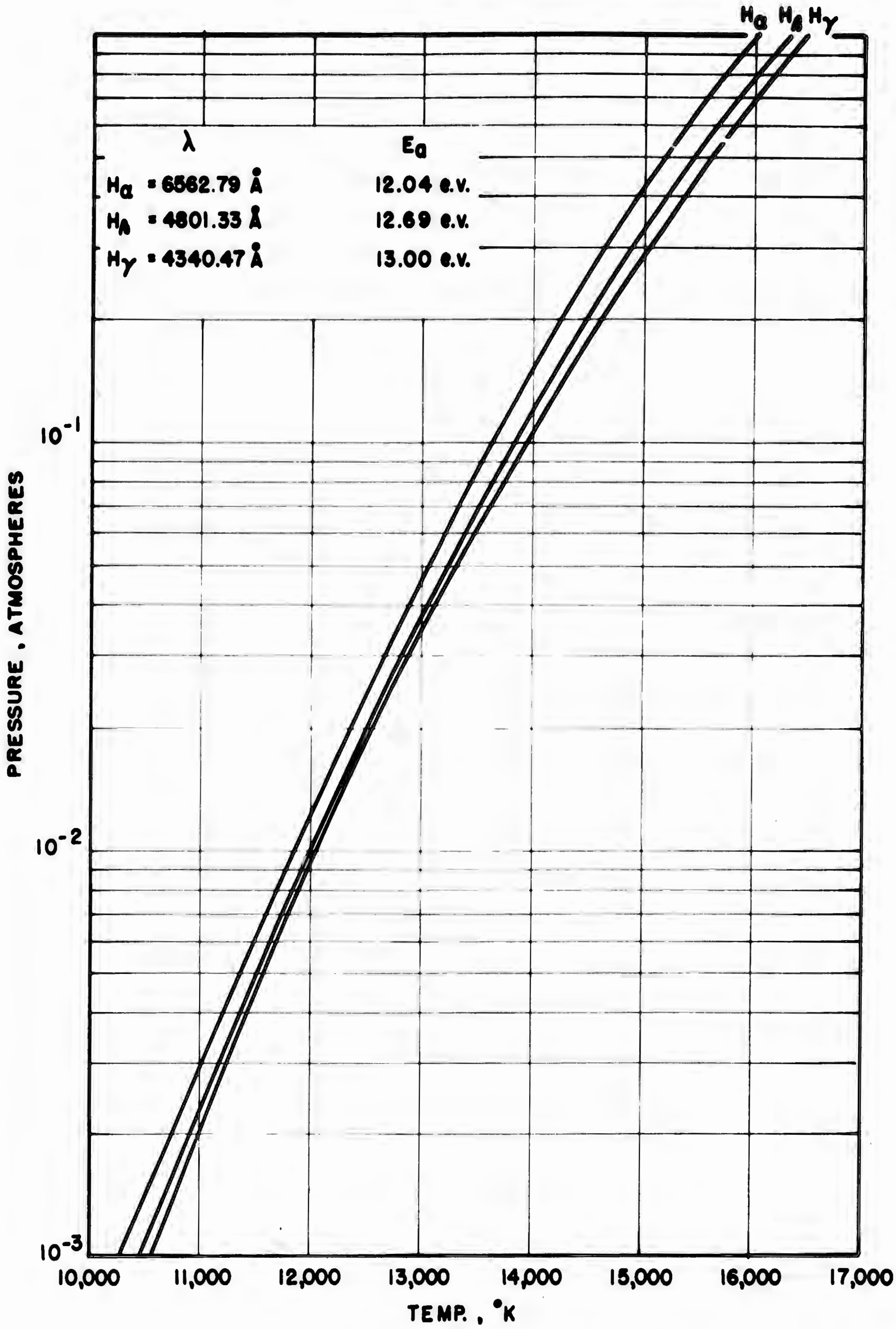


FIG. 7 - THE TEMPERATURE FOR PEAK LINE INTENSITY AS A FUNCTION OF AMBIENT PRESSURE (HYDROGEN)

WOUND HEALING, COLLAGEN, AND HUMORAL GROWTH PROMOTING AGENTS

H. ROSEN, C. W. BERARD, E. F. GEEVER, and S. M. LEVENSON
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C.

The healing of a wound is a dynamic process, involving mechanisms, both general and unique, which manifest themselves through a spectacular progression of biochemical and cellular changes during the healing period. One of the most striking of these phenomena is the creation of collagen, a substance absolutely vital to the healing process; this material has constituted our major field of interest in the past few years. We have attempted to approach the following questions: What is collagen? How is it synthesized and degraded by the living organism? What is the role of ascorbic acid in these processes, and what is the relationship of collagen to the "ground substance", that material in which the fibers lie buried, and which presumably confers on them dimensional stability?

Collagen is the fundamental structural material of the animal kingdom. One finds it listed in the biochemistry textbooks as a "scleroprotein", that is, one of those proteins including keratin which will dissolve in no known salt solution. It has been classically distinguished from the other members of its family by its unique ability to dissolve in hot water. It is now known that collagen will, to a certain extent, dissolve in salt solutions, and that this "salt soluble" collagen is itself heterogeneous. Thus, there exists a number of different collagen fractions derived from the parent protein macromolecule. This parent macromolecule is analagous to the more familiar protein-macromolecular fibers, such as silk or wool.

Collagen is apparently synthesized within the fibroblast and extruded into the extracellular spaces, where it develops by an as yet unknown process or processes into the giant, macromolecular, banded strands called collagen. Transverse striations are characteristic of collagen, and lie about 640 Å apart in native collagen. The amino acids hydroxyproline and hydroylysine constitute, apparently, another characteristic property unique to collagen. The collagen macromolecule has been considered to be a very long one composed of

the three inextricably intertwined strands, each with repeating subunits of the same structure. However, it has thus far been impossible to isolate from the huge strands of the material a unit to which the term "molecule" could be properly applied. The term "tropocollagen" has been assigned to a theoretical building block which supposedly exists in native collagen (1), and fragments of collagen have been visualized by the ultramicroscope, which are thought to correspond to "tropocollagen" (2). This "unit" however has never been isolated and remains hypothetical.

It has been known since before 1870 (3) that collagen swells in dilute acetic acid, loses its fibrillar appearance, and eventually dissolves, in part, to yield a viscous solution. Within the last few years American workers have postulated that dissolved collagen should contain more than one species (4), and Russian investigators have fractionated dissolved collagen into two components (5). Attempts at detailed examination of the chemical and physical characteristics of these components have been hindered by the inability to isolate them adequately. The techniques (6) used to fractionate those substances known commercially as "gelatin"--collagens more or less drastically treated with acids or alkalies--do not seem satisfactory for this purpose. Ion exchange chromatography (7) requiring sodium hydroxide is drastic for the study of primary structure; as for other methods, Orekhovich and Shpikiter (8) found alcohol precipitation (9) and coacervation (10) unsatisfactory for the separation of the components demonstrated by ultracentrifugation of solutions of collagen treated with hydrogen bond-breaking agents. We have developed a method which we find satisfactory and have succeeded in chromatographing rat tail tendon collagen into at least four major components. The method involves gradient elution chromatography on warmed carboxymethyl cellulose columns, of acetic acid solubilized collagen, under mild conditions of pH and low salt concentrations (11). Rat tail tendons were cut into 1-cm pieces, directly immersed in 100 ml of chilled 0.16 or 0.2 M sodium chloride and shaken slowly for 18 hours at 4°. This sodium chloride extract was discarded, and the tendons washed with 5 portions of chilled distilled water; each rinse was 50 ml and lasted 5 to 10 seconds. After the addition of 35 ml of 0.1M acetic acid to 240 mg of wet tendon, the mixture was allowed to stand in a tightly stoppered Erlenmeyer flask at various temperatures and for various times. The mixture was centrifuged and filtered. The result was a clear viscous solution of collagen which was usually chromatographed immediately, although no changes were detected in chromatographic patterns of solutions left 4 months at $4^{\circ}C$. The carboxymethyl cellulose was prepared from Whatman 200 mesh wood pulp cellulose by the method of Peterson and Sober (12). Columns were 0.9 cm in diameter and jacketed for water circulation allowing precise control of the 40° temperature required for the chromatography. After the column had packed by gravity, it was washed with the starting buffer until the effluent was no longer basic, and then with about 4 ml of distilled water, and allowed to become dry at the top.

A sample (1.5 ml) of the 0.1 M acetic acid collagen solution, containing about 0.5 mg of nitrogen, was applied to the heated column (40°) at this time. This was the first time that the collagen solution was subjected to a temperature higher than 20°-24°. The sample was allowed to run into the adsorbent, covered with 4 ml of initial buffer, and the elution gradient process was started. The increase of sodium ion in the elution gradient process was nearly linear for most of its course, starting at 0.05 M sodium ion, and reaching a maximum of 0.165 M sodium ion at the end of the chromatographic process. Two-milliliter fractions were collected at a flow rate of 28 ml per hour; each run took about 7 hours; recovery of collagen from the columns in one experiment in which all the fractions were combined, with Kjeldahl nitrogen as an index, was 104%. Amino acid analyses, with the exception of hydroxyproline, were performed by the chromatographic method of Moore and Stein (13) on columns of Amberlite IR-120, and determined quantitatively by the method of Rosen (14).

We found by this procedure that rat tail tendon collagen, dissolved in acetic acid actually consists of four major parts (Fig. 1). This acetic acid collagen solution, however, shows only one component in the ultracentrifuge. Three of the fractions we isolated are high molecular weight, gelatin-like components, each of which has the singular amino acid composition characteristic of the parent collagen, that is, high glycine, proline and hydroxyproline concentrations. The other fraction, however, is a mixture of small peptides, almost quantitatively dialyzable, containing much glycine and proline but little hydroxyproline, in distinct contrast to the other fractions.

In a number of runs, we carried out the entire procedure sterilely up to the chromatography itself; these chromatograms were identical to the ones in which no sterile precautions were taken. Bacterial processes, then, seem to play no role in the collagen fractionation achieved by our procedure.

The process by which the fractions are created seems to be an "all or none" one, since experiments showed that the characteristic four-fraction pattern appeared at the earliest time we could manage to complete the steps preliminary to chromatography (10 minutes). Such an "all or none" process was also suggested by Boedtker and Doty (15) for the breakdown of dissolved ichthyocol in warm citrate buffer. It seems reasonably certain that hydrolysis played no significant part in the production of the rat tail collagen components, since our chromatographic procedure required only 7 hours and the peptide fraction was completely eluted in as little as 10 minutes. The solutions retained their chromatographic properties after being kept at 25° C for 148 hours and changed only slightly when kept at 40° up to 48 hours. After 48 hours at 40° C the chromatograms were smeared; this may have been due to hydrolysis. Boedtker and Doty found what seems

to have been a similar process of hydrolysis for ichthyocol dissolved in pH 3.7 citrate buffer, the viscosity of which fell slowly with time when the solution was kept for over 40 hours at 36° C.

With respect to the chromatographic separation of the last three components, we have no evidence for the nature of the bonds cleaved, nor for the structural features upon which the fractionation depends. These features may be physical, for example, molecular weights, or chemical, for example, differences in frequency of occurrence of amide groups. Preliminary analysis indicates that the three last fractions differ slightly in amino acid composition. The unique amino acid composition of the peptide fraction has already been mentioned. Thus, none of these collagen components could correspond to "tropocollagen" or to any other single, repetitive collagen building block. Gallop, et al. (16) have recently suggested the existence of ester-like bonds within ichthyocol which give rise to nondialyzable components when split with hydroxylamine or hydrazine in such a way that peptide bonds are not disrupted. It is possible that such bonds are also involved in the phenomena which we are discussing.

The first fraction obtained from the relatively dilute acetic acid solutions is interesting in that it contains so few amino acids, and that its hydroxyproline content is so low. It might be imagined that this peptide fraction is a contaminant. This is unlikely, since (a) the tendons were washed prior to analysis with 0.2 M sodium chloride at 4° C for 24 hours, treatment which might be expected to dissolve any clinging contaminants and (b) the singular amino acid distribution of these peptides does not correspond to any known mammalian protein, including "classical" collagen. This component, then, is evidence of the existence of discrete atypical areas within the collagen molecule; its small hydroxyproline but great glycine and proline content gives us reason to postulate that it may represent an early stage of collagen formation and may, in fact, be the intermediate through which proline is converted to hydroxyproline. This postulate is based in part on the observations of Stetten (17), that the formation of the hydroxyproline of collagen is unusual since it is formed from proline at or near the site of its introduction into the macromolecule. Free exogenous hydroxyproline is not utilized in collagen formation.

To test this hypothesis and to determine the physiological significance of the other three collagen fractions as well, we performed tracer experiments with C^{14} glycine. Collagen was obtained from the tail tendons of young growing rats at various times (3 hours to 6 weeks) after they had been injected subcutaneously with the labeled amino acid. The collagen samples were solubilized in 0.1 M acetic acid. These solutions were chromatographed on carboxymethylcellulose at 40° C as already described. The four components were identified and measured in the eluate by the Lowry method and by liquid scintillation counting. Sequential analyses showed that while each of the four fractions attained its maximum radioactivity

twelve hours after injection, two of the fractions attained strikingly higher specific activities than the others, and they lost these activities more rapidly also (Fig. 2, 3). The two components showing the unusually high turnover rates were the first fraction (peptides) and the second fraction (first of the high-molecular weight components).

Orekhovich, et al. reported recently an experiment in which carboxyl-labeled glycine-C^{14} was injected into rats (5). Skin collagen solubilized with citrate buffers was subsequently fractionated into two components ("α" and "β") by ammonium sulfate precipitation of the partially urea-denatured solution of collagen. They found that their α fraction had attained a specific activity about three times that of the β component 12-15 hours after injection. It will be recalled that our own two active components of rat tail tendon collagen were twice as radioactive as our two less active ones 12 hours after injection. It seems possible, therefore, that our second fraction (first of the high molecular weight components) may correspond to the α fraction of the Russian workers, and that their β fraction is a composite of our fractions 3 and 4. Orekhovich and his colleagues did not isolate a low molecular weight component.

It is clear, then, that not only is collagen inhomogeneous physically, but that the fractionated components we have found were formed and metabolized at different rates. The unusual amino acid composition of the peptide fraction, together with its low molecular weight and very high metabolic activity, supports our hypothesis that it represents long sought for collagen precursor.

We have made a long series of studies of the peptide fraction with highly automated equipment designed in this laboratory for this particular function, but applicable to studies of amino acids and peptides in general. The equipment makes use of ion exchange and a gradient elution scheme devised by Piez and Morris (18) for use with automatic flow and recording devices. We have simplified the procedure by eliminating the use of highly unstable ninhydrin solutions, and by eliminating troublesome mechanical deficiencies. Our procedure uses a clog-free, stable ninhydrin solution prepared in air and light and used with no protection necessary from these elements. This is an adaptation of another method, also developed by us (14). Our experiments have led us to the conclusion that there may be as many as fourteen separate peptides within the small molecular weight fraction. We are determining the amino acid composition of these peptides. This, together with measurements of the specific metabolic activity of each, should delineate more exactly the biological role of these peptides and make it possible to reconstruct some aspects of the biosynthetic processes which give rise to them.

It is our thought that ascorbic acid may exert its influence on collagen formation at this point of the synthetic process. Although it has been known for over 200 years that wounds do not heal in

patients with scurvy, the mechanism whereby vitamin C affects collagen synthesis is unknown. Our working hypothesis is that ascorbic acid influences collagen metabolism by regulating the formation of the peptide precursors which we have mentioned, and the rate at which their many proline residues are converted to hydroxyproline. If this is so, it provides one of the few examples, or only example, of a demonstrable peptide precursor to an eventual protein in mammals. Thus, the "template" mechanism, currently postulated as the mechanism of protein synthesis, might not be a universal one, after all, since in its classical form it requires the synthesis of protein whole, not in sections.

Why does a wound begin to heal, what sustains the healing process, and what stops it? From time to time, evidence has been presented by various investigators that humoral factors, perhaps originating locally, are involved in these processes. But the evidence has been inconclusive. During the past few years, working with wound fluid provided to us by Dr. John S. Schilling of the University of Oklahoma Medical School, we have shown the existence of an ultrafilterable growth accelerating substance, probably peptide, during the rapidly "healing" phase after subcutaneous implantation of cylinders of stainless steel wire mesh.

When these cylinders are implanted, they are quickly filled with serum and fibrin and contain varying numbers of red blood cells, macrophages, and leukocytes. There then evolves over the next several weeks the full pattern of proliferating granulation tissue, complete with fibroblasts, capillaries, and fiber formation. The picture resembles closely that seen in healing wounds. The fluid which bathes the interfaces of the proliferating fibro-collagenous tissues may be obtained by aspiration from the interior of the cylinders. We call this "wound fluid", and have obtained it at varying periods during the time-course of fibro-collagenous incapsulation and internal obliteration of the space of the cylinder. As many as 20 cylinders were implanted subcutaneously in dogs and from 0.5 to 2.0 ml of wound fluid were collected from each cylinder. We were able to obtain in this way large amounts of the fluid for a variety of analytical procedures.

We found first that very small quantities of an ultrafiltrate of the wound fluid obtained during the first week following implantation of the wire-mesh cylinders accelerated dramatically the growth of L. casei in a chemically-defined amino acid medium (Fig. 4). Ultrafiltrates of plasma taken from the dogs at the same time the wound fluids were obtained had no such effect. We then traced the appearance, accumulation, and disappearance of this factor in the wound fluid as healing progressed (Fig. 5). It is clear that the growth-promoting effect is highest early and then declines progressively though even at three weeks after implantation the effect is significantly higher than that of plasma ultrafiltrates obtained at

this time.

We have partially isolated and partially characterized this growth accelerator. It is nonprotein, somewhat basic, water and n-butanol soluble, but insoluble in ethyl acetate and other organic solvents. It loses 50% of its activity after heating for 4 hours at 100° C. It is not a lipid, amino acid, simple sugar, glucosamine, kinetin, or a complex lipid. Our current postulate is that it is peptide in nature.

Our working thesis is that this substance speeds the transport of amino acids and other critical metabolites across the cell membrane and thereby speeds protein synthesis. We also feel that its significance goes beyond that of accelerating bacterial growth alone and that it may be fundamental for growth in general, wound healing and organ regeneration. We have adduced some confirmatory evidence for this hypothesis by our finding that this wound fluid ultrafiltrate accelerates the growth of mammalian (mouse fibroblast) as well as bacterial cells. Table I shows results of a typical study of this type, carried out for us by Dr. Andre D. Glinos of the Growth Physiology Laboratory of our Division. The "control" was a sample of plasma ultrafiltrate isolated from the dog when the wound fluid was taken. The wound fluid ultrafiltrate effected an almost two-fold increase in rate of cell division.

Finally, we have recently undertaken a study of the three-dimensional configuration of collagen, its *in vivo* stability and its relation to the ground substance, by the introduction of heavy water into the diet of animals. It is known that heavy water alters the stability of proteins *in vitro*. We became interested in heavy water as a tool for the study of collagen when we used it in an *in vitro* study of the viscosity characteristics of collagen. Rat tail tendon collagen was dissolved first in various concentrations of acetic acid in ordinary water. These solutions were then held at or near the "melt" temperature of collagen, about 39° C. This is the temperature at which such solutions very rapidly lose their viscosity, due to thermal rupture of the hydrogen bonds which hold the macromolecules in their typical elongated helical configuration. Breaking of these bonds results in the conversion of the collagen to a "random coil" shape, much more nearly spherical, and therefore of very low viscosity. We found, to our surprise, that when rat tail tendon collagen was dissolved in acidic heavy water, the "melt temperature" was raised by an appreciable amount. Solutions of collagen which would normally uncoil in an hour at 39° C would now take several days to undergo the same conversion. Heavy water, in essence, acted as a "helix" protector, probably by hydrogen bond reinforcement

We have now carried out experiments with heavy water *in vivo* where the stability of collagen depends not only on its own inherent characteristics, but also on its relationship to the relatively acidic

ground substance. Rats were subjected to our standard 5 cm dorsal skin incisions (19) plus the abdominal subcutaneous implantation of polyvinyl sponges (20) (Fig. 6a, 6b.) Granulation tissue consisting of newly formed capillaries, fibroblasts, ground substance and collagen fibers develops in the plastic sponges. The process is similar to that which occurs in the wire mesh cylinders and in the healing incisions. Excision of the sponge at any chosen time provides a specimen of pure reparative collagen.

The incisions and sponges were examined histologically using a variety of staining techniques. In addition, the tensile strengths of the incisions were measured by a technique devised in our laboratory and the sponges with their contained granulation tissue were subjected to various chemical analyses. Included among the latter analyses were measurements of hydroxyproline as an index of collagen. We found that although the histologic appearance and hydroxyproline content of the sponges of the deuterated animals were normal, the tensile strength of their wounds was about 40% lower than that of the controls (Table II). Surprisingly, when both sets of wounds were soaked in formaldehyde, the deuterated ones gained relatively more strength, indicating that the formaldehyde had corrected, probably by methylene bridges, a basic collagen structural deficiency in the wound of the deuterated animal. Histological examination of the deuterated wounds showed the collagen less organized and finer in structure than the normal. In addition, skin and tail tendons of these animals were more susceptible to attack by collagenase in vitro and were relatively less soluble in neutral salt solution. All these phenomena indicate that heavy water produced an in vivo abnormality of both reparative and developmental collagen, or of its interaction with the ground substance, or both.

In summary, we have applied to the problem of wound healing a wide variety of physical, biochemical, surgical, and histological techniques. We have succeeded in showing that dissolved collagen separates into discrete fractions, and that one of these is a peptide complex with an atypical amino acid content; we have separated these peptides into 14 fractions by ion exchange chromatography; we have shown that these fractions are metabolically distinct and postulate that they may represent stages in collagen development; it is possible that ascorbic acid, acting on the peptides, converts them from predominantly proline types to predominantly hydroxyproline types, creating hydrogen bonding sites by which the peptides are tied into the main part of the macromolecule. We have produced an apparently unique collagen defect by forcing mammalian collagen to develop in a milieu of heavy water. We hope, through further study of this defect, to discover basic facts concerning the tertiary structure of naturally occurring collagen, and how it interacts with the mucopolysaccharides of the ground substance. Finally, we have found, in mammalian wound fluid, a substance which accelerates bacterial growth and the division of mammalian fibroblasts, and which may act in vivo, as a "wound hormone."

Table I. Effect of Wound Fluid Ultrafiltrate on the Growth of L Strain Cells in the Absence of Serum Protein

Material Added to Basal Medium	Initial No. of Cells x 10^3	Final No. of Cells x 10^3	Growth Ratio = Final No. Cells / Initial No. Cells
None (Control)	300	448	1.5
Wound Fluid Ultrafiltrate 3%	300	717	2.4
Wound Fluid Ultrafiltrate 6%	300	805	2.7
Wound Fluid Ultrafiltrate 12%	300	805	2.7

Table II

Fresh Wounds		Formaldehyde Fixed	
D_2O	Control	D_2O	Control
3.50	5.60	6.80	7.20
4.00	6.60	10.60	8.00
3.55	4.85	8.65	12.00
3.65	5.95	5.85	11.75
4.65	3.95	7.00	11.00
5.15	4.50	6.05	11.05
3.00	5.25	9.10	7.95
3.75	5.80	8.50	8.00
3.50	4.50	9.05	5.10
4.45	5.85	8.00	7.80
4.70	5.90	5.20	6.05
4.90	6.55	6.85	6.30
3.40	4.10	7.50	6.05
4.00	5.20	6.00	5.40
3.25	7.65	6.75	7.40
3.50	8.00	8.30	7.50
3.95 (Ave.)	5.64 (Ave.)	7.49 (Ave.)	8.04 (Ave.)
S.D. = $\pm$ 0.65	S.D. = $\pm$ 1.16	S.D. = $\pm$ 1.46	S.D. = $\pm$2.24
S.E. = $\pm$ 0.16	S.E. = $\pm$ 0.29	S.E. = $\pm$ 0.37	S.E. = $\pm$0.56
$P < 0.001$		$P > 0.1$	

REFERENCES

1. Schmitt, F. O., Gross, J., and Highberger, J. H., Exp. Cell Rsch. Suppl 3: 326, 1955.

2. Hall, C. E., Proc. Natl. Acad. Sci., U. S., 42: 801, 1956.

3. Gillette, E. P., Germer Bailliere, Libraire Editeur, Paris, 1872.

4. Doty, P., and Nishihara, T., In G. Stainsby (Editor), RECENT ADVANCES IN GELATIN AND GLUE RESEARCH, Pergamon Press, Ltd., London, 1958, p.92.

5. Orekhovich, V., Shpikiter, V., Kasakova, O., and Mazourov, V., Arch. Biochem. Biophys., 85: 554, 1959.

6. Pouradier, J., and Venet, A. M., J. chim. phys., 47: 11, 1950.

7. Russell, G., Nature, 181: 102, 1958.

8. Orekhovich, V. N., and Shpikiter, V. O., In G. Stainsby (Editor), RECENT ADVANCES IN GELATIN AND GLUE RESEARCH, Pergamon Press, Ltd., London, 1958, p.87.

9. Dervichian, D. G., and Van Den Berg, M. J., Bull. soc. de chim. biol., 30: 559, 1948.

10. Stainsby, G., Faraday Soc., 18: 288, 1954.

11. Kessler, A., Rosen, H., and Levenson, S. M., J. Biol. Chem., 235: 989, 1960.

12. Peterson, E. A., and Sober, H. A., J. Am. Chem. Soc., 78: 751,1956

13. Moore, S., Spackman, D. H., and Stein, W. H., Anal. Chem., 30: 1185, 1958.

14. Rosen, H., Arch. Biochem. Biophys., 67: 10, 1957.

15. Boedtker, H., and Doty, P., J. Am. Chem. Soc., 78: 4267, 1956.

16. Gallop, P. M., Seifter, S., and Meilman, E., Nature, 183: 1659,1959.

17. Stetten, M. R., J. Biol. Chem., 181: 31, 1949.

18. Piez, K. A., and Morris, L., Anal. Biochem., 1: 857, 1959.

19. Geever, E. F., and Levenson, S. M., A.M.A. Arch. Ophth. 63: 812, 1960.

20. Boucek, R. J., and Noble, N. L., A.M.A. Arch. Path., 59: 533,1955.

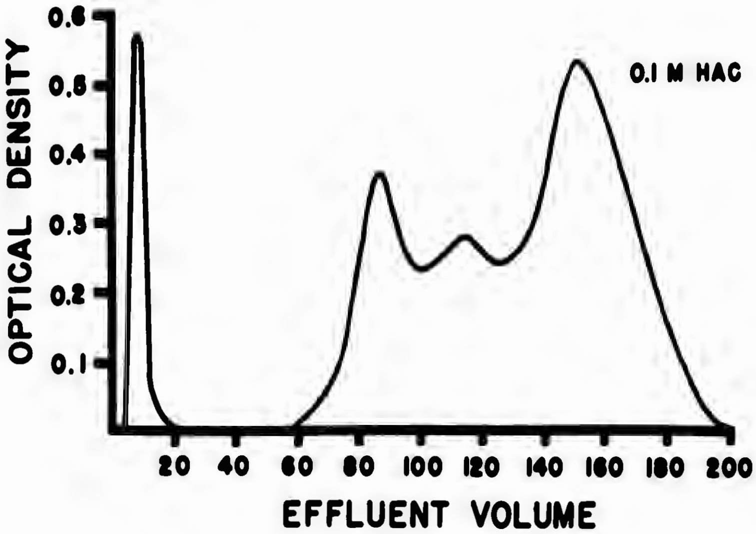


Fig. 1 -- Chromatographic pattern of acetic acid solubilized rat tail collagen.

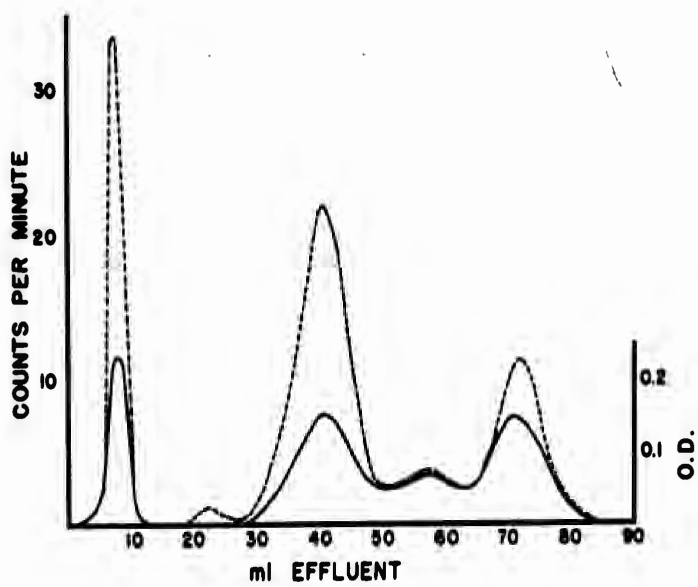


Fig. 2 -- Chromatographic and radioactivity patterns; rats injected with C^{14} glycine.

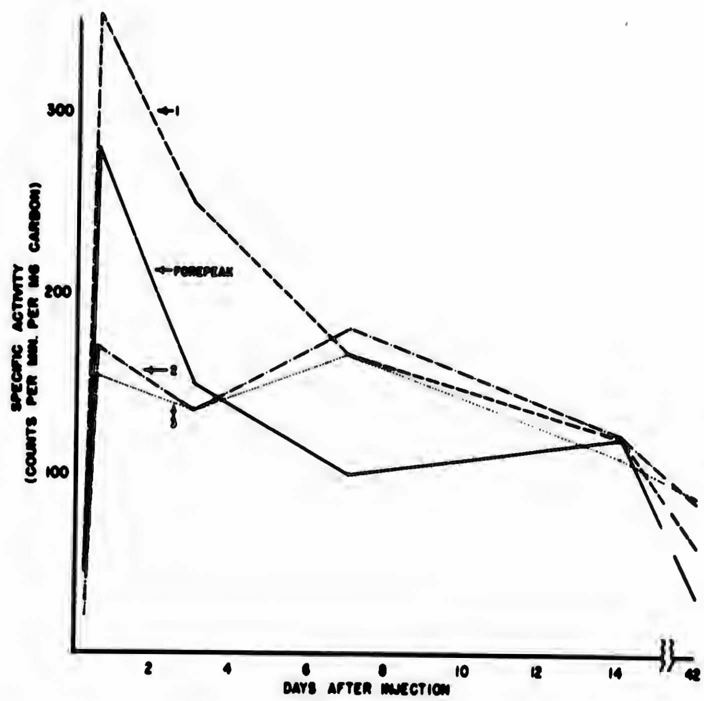


Fig. 3 -- Rate of appearance and disappearance of C^{14} activity from fractions of tail collagen.

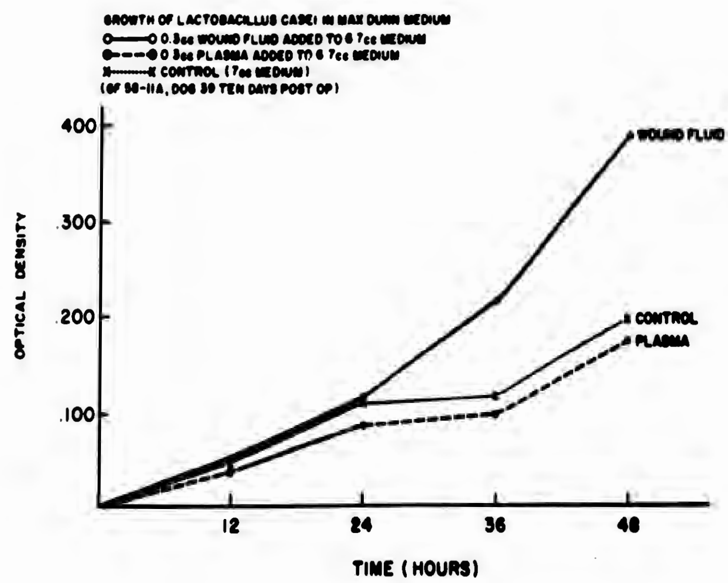


Fig. 4 -- Acceleration of L. casei growth by dog wound fluid ultrafiltrate.

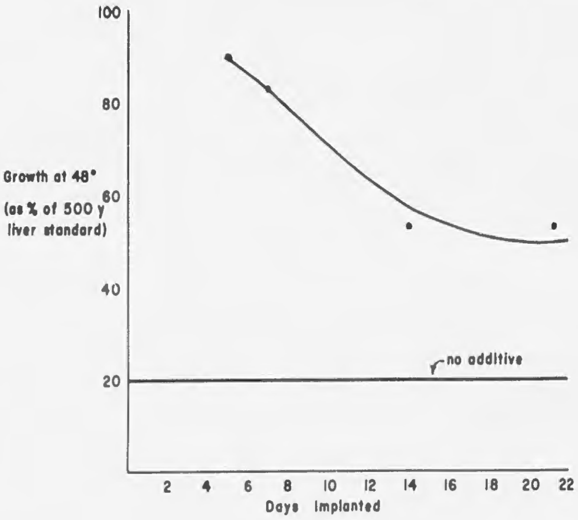


Fig. 5 -- Effect of implantation period on growth accelerating properties of wound fluid.

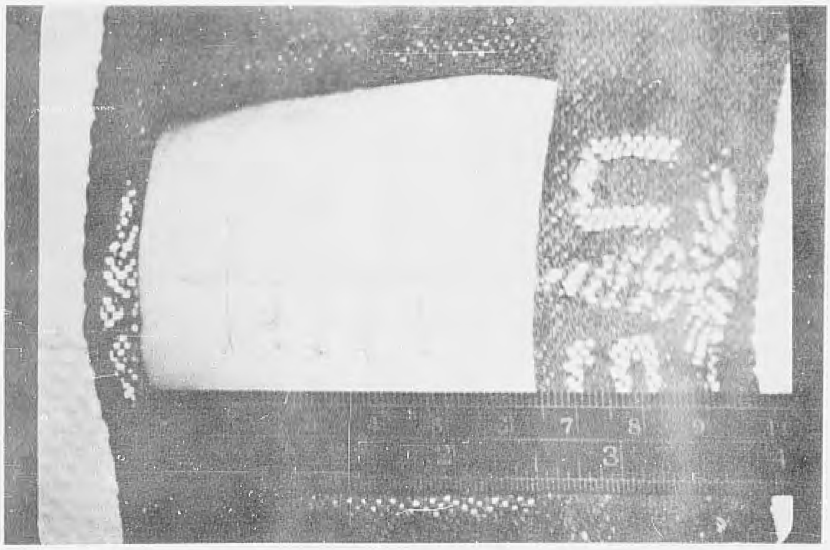


Fig. 6A -- Five cm. dorsal skin incision in the rat.

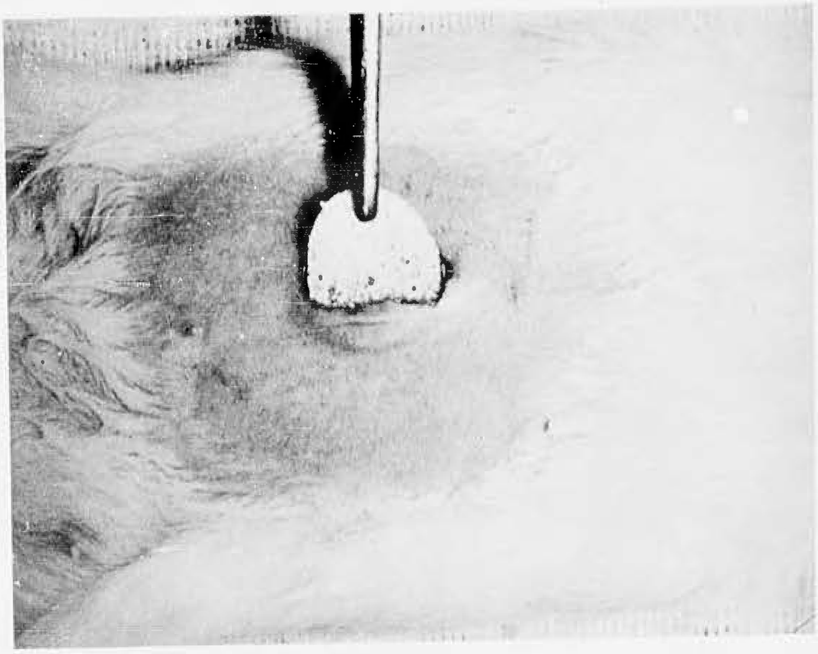


Fig. 6B -- Implantation of 25 mg polyvinyl sponge into the rat.

TECHNIQUES FOR OPTIMIZING IMAGE INTERPRETER PERFORMANCE

ROBERT SADACCA
U. S. ARMY PERSONNEL RESEARCH OFFICE
WASHINGTON 25, D. C.

From its image interpreters the Army needs information which is timely, relevant, complete, and accurate. Until recently, the Army's image interpreter was confronted only with the problem of interpreting relatively large-scale, conventional black-and-white photographs. Today he is also required to interpret small-scale and degraded photographs, as well as radar, infra-red, and TV imagery. All the military services are currently increasing both their capacity to obtain imagery and their capacity to process imagery at a rapid rate. This increased capacity and variety of image sources have inevitably increased the quality range of images. But good or poor, the imagery obtained through these media is ultimately placed before the image interpreter who is asked to extract information useful for important military decisions.

Existing knowledge of the basic psychological factors operating in the interpretation of real imagery is severely limited. Previous research effort in aerial surveillance intelligence systems has tended to concentrate on improving the vehicles or platforms from which the aerial imagery is collected and also on improving the physical characteristics or quality of the obtained imagery. In response to a requirement generated by the Assistant Chief of Staff for Intelligence the research program of the U. S. Army Personnel Research Office concentrates on the man within the system to determine how the skills, knowledges, and techniques which he brings to the job impinge upon his performance and the performance of the system as a whole. Thus, we are concerned with image interpreters' interests, aptitudes, and experiences, with the background information and instructions they receive, with the viewing techniques and the scan procedures they employ, with the clues that they selectively use in interpreting particular photographs, and in their use of references and other aids--target keys, bombing encyclopedia, electronic comparators, stereoscopes, tube magnifiers. It was recognized from the beginning that the performance of image interpreters would not be perfect; this is,

of course, the case for all individuals making complex judgments and inferences. Because of the criticality of the interpreter in the aerial surveillance cycle, his job was the first selected for intensive study.

In focusing on the image interpreter, we are attempting to identify the significant sources of variance underlying observed differences in performance with the aim of ultimately optimizing the role of the interpreter in the system. Investigation of men within the aerial reconnaissance system is complicated by the fact that they do not come into the system ranked according to measured abilities or even labeled as possessing a certain amount of a given characteristic. Individual differences among image interpreters must be measured and described and functional relationships among variables dependent upon individual differences must be obtained. In research to determine effective interpreter techniques, these individual differences must be controlled as well. We have attempted to equate the image interpreters working under various experimental conditions through random assignment of matched interpreters. However, within condition performance variance has always been high; and, owing to the small size of available samples, statistical significance has not always been achieved even where data trends and practical considerations point to real differences among experimental techniques or conditions. The very fact that interpreters do vary so widely in performance points to the necessity of thoroughly studying their role within the system before the system as a whole can be optimized.

There are two important aspects of image interpreter performance which can be measured and quantified. The first is the number of correct identifications of militarily significant objects or targets made in the available time period. In an operational situation, most--if not all--of the important enemy objects actually on the imagery must be correctly located and identified. In addition, the number of erroneous identifications made by interpreters can be determined. The number of these errors must, of course, be minimized. This second aspect may not at first appear to be as important as the first. However, it must be remembered that aerial imagery is often of very small scale and that quality is not always the best. While the number of actual targets on any one image limits the number of correct identifications that can be made, the number of misidentifications that can be made is not so limited.

Errors made by interpreters subdivide into two types. Interpreters can misidentify militarily important objects--what is really a bunker, for example, can be identified as a tank; or a chemical plant can be identified as a steel plant. This "misidentification-of-significant-object" type of error is more common in reports for strategic imagery, where the objects tend to be relatively large and well defined, than for tactical imagery. Interpreters can also make what we term "inventive" errors, that is, they can identify insignificant

features of the photographs--road tonal differences, for instance--as significant targets. The tendency to make "inventive" errors is more common in the interpretation of tactical photographs, in which the targets are relatively small and dispersed and in which there are many variations in shape and tone.

Conflict arises when attempts are made to increase the number of correct target identifications and at the same time reduce the number of erroneous identifications. Work methods which tend to maximize the number of correct identifications often tend to yield more wrong identifications. Similarly, procedures that tend to minimize the number of wrong identifications often tend to yield fewer correct identifications.

It might be thought that the way out of this dilemma is to concentrate on maximizing the ratio of right identifications to total identifications made. This would in effect maximize accuracy, or percent right. But would this completely solve the problem? Can we say, for example, that a method which yields 4 actual military targets and 1 "wrong" target--an accuracy of 80%--is better in all circumstances than a method which yields 30 "right" targets and 20 "wrong" targets--an accuracy of only 60%. Much obviously depends upon the intelligence requirements of the military situation: In an effort to get the maximum number of actual targets identified, a commander may tolerate many misidentifications. If, on the other hand, he must be very sure of the targets to be attacked, he may insist upon positive identification of a few targets, even though others remain unidentified. As a corollary, the optimal work methods for interpreters may vary with the intelligence requirements of the military situation.

THE PERFORMANCE MEASURES

Performance measures were used in all our studies to gauge the effect of various work procedures on the speed, accuracy, and completeness of image interpretation. Before presenting some of the results obtained, a brief description of how we developed our performance measures may be in order. The performance measures reproduce, as far as possible, typical real-life situations which confront the image interpreter. The imagery derives from actual tactical and strategic air reconnaissance missions. Some photographs, taken during war time, show enemy-held positions. The objects the image interpreter was directed to look for reflected the intelligence requirements of the military situation at the time of the air reconnaissance mission. As in an operational situation, the interpreter was provided with maps of the relevant areas, sortie plot overlays, and general and specific information about the location and disposition of friendly and enemy forces. Also, as in the operational situation, the image interpreter was allowed to examine the imagery presented to him in any way he desired and to use any instruments in his kit.

Photographs were printed from the original mission negatives. Interpreters examined have agreed that the print quality was good-to-excellent, and in general better than they are accustomed to working with. The scale of the photos varied from 1:4,000 to 1:8,000 for the tactical interpretation exercises and from 1:10,000 to 1:15,000 for the strategic interpretation exercises. The scales were considered adequate for identifying the objects for which the interpreters were required to search.

EXPERIMENTAL WORK METHODS

This paper describes four experimental techniques that were studied in an effort to gain insight into how interpreter performance might be improved. Procedures were varied in order to estimate how interpreter performance was affected by:

(1) the interpreter's confidence in his identifications
(2) the length of time spent in viewing
(3) repeated exposure to a given set of imagery
(4) different team work methods

The techniques were tried out on image interpreters just before their graduation from the Intelligence School. Comparisons were made between or among groups of interpreters matched on the basis of their final course grades and scores on several aptitude tests administered prior to formal course training. In examining these results it must be borne in mind that the interpreters in our samples had no on-the-job experience. In connection with this, it should be stressed that the main purpose in conducting these experimental studies was to determine under controlled conditions, techniques and procedures that might be useful in an operational setting. Until research is extended to the operational setting, these findings can only be considered as suggestive. Further, it is not possible to directly assess the contribution to intelligence estimates that image interpreters make on the basis of performance of the graduating interpreters used in these experiments.

THE INTERPRETER'S CONFIDENCE IN HIS IDENTIFICATIONS

One of the first studies conducted was to determine whether the degree of confidence interpreters expressed in their identifications was related to the accuracy of their identifications. In the operational situation an interpreter normally indicates the degree of confidence or certitude he has in an identification by one of the following qualifiers: "Positive", "Probable", or "Possible". The same procedure was followed in the performance measures. Interpreters were told to place a <u>1</u> next to those identifications they were positive about, a <u>2</u> next to probable identifications, and a <u>3</u> next to possible identifications. Table 1 shows the mean confidence level for right and wrong identifications found in five performance measures

for several groups of interpreters. The mean confidence level for right identifications was consistently higher than for wrong identifications. This finding has practical implications. If systematic advantage can be taken of accompanying expressions of confidence, the amount of wrong information produced by interpreters may be substantially reduced.

As a first step in this direction, the effect of increasing the customary 3-point confidence scale used by interpreters to a 5-point scale was explored. Two groups of matched interpreters were established. One group used the regular 3-point scale. The second group used a 5-point scale with the following confidence levels: Highly Positive, Positive, Highly Probable, Probable, and Possible. Two performance measures, one involving tactical and the other strategic imagery, were administered under identical conditions to both groups of interpreters.

With the 3-point scale, percent accuracy fell off sharply for identifications interpreters felt were only probable--from 32 to 13% in the tactical photos and from 71 to 45% in the strategic photos (see Table 2). The 5-point scale seemed to achieve a more even distribution of accuracy percentages, although the small accuracy differences between the lower levels of confidence left much to be desired. The higher accuracy rates obtained for identifications which interpreters using the 5-point scale considered highly positive point to the possibility that erroneous identifications might be reduced by accepting only identifications of which interpreters are highly confident. However, such a procedure would lose the benefit of a large number of correct identifications which received lower confidence ratings. Although the number of wrong identifications might be minimized, the number of correct identifications would also be reduced. Further research, trying out various scales with different interval steps, is needed before interpreters' feelings of certitude concerning their identifications can be best put to use operationally.

LENGTH OF VIEWING TIME

The objective of another study was to determine whether the rate of information extraction varied systematically over time. The cumulative number of identifications made after each five-minute period of viewing time was computed for right, wrong, and total number of identifications for several performance measures. These values, when plotted, resulted in fairly uniform curves. Figure 1 shows the curves for a typical example--a tactical measure administered with a 30-minute time limit. Notice that the average number of new identifications scored right tapered off much sooner than did identifications that were wrong. Although the interpreters were still responding at the end of the 30-minute time period, the later information contained a higher proportion of wrong identifications than did information produced earlier in the viewing period.

The effect is seen more clearly in Figure 2 which shows accuracy plotted as a function of time for the same performance measure. The average accuracy of the interpreters fell off the longer they examined the imagery. Evidently, time does not necessarily work to the advantage of the image interpreter. After examining imagery for a while, interpreters may begin to respond to doubtful cues, cues that are on the threshold of resolution. Also, the very fact that interpreters are asked to locate specific targets may suggest the presence of the targets.

The problem of how much time interpreters should be allowed to examine imagery needs to be further investigated. Here again, the problem of optimization arises. If an interpreter's time is cut short so that he will not be induced to make doubtful identifications, some correct identifications will probably be lost. New techniques need to be developed for establishing the viewing time at which the desired balance is achieved between number of right and wrong identifications at a high level of accuracy.

PERFORMANCE ON RE-EXAMINATION OF IMAGERY

In another study the usefulness of having interpreters re-examine imagery they had interpreted once before was explored. It was hypothesized that, on a second exposure to the imagery, interpreters would make more correct identifications and fewer errors than they did when they first interpreted the imagery. Two performance measures, one containing tactical photos and the other strategic photos, were administered to a group of 29 interpreters. The same performance measures were readministered two and one-half days later. Results were in some ways disappointing. Mean number of right identifications and mean accuracy scores on the second administration were not significantly different from the comparable scores obtained on the first administration, although slightly higher. Mean wrong scores on the second administration were also up slightly from the first administration.

However, further analysis of the data revealed one significant difference ($P < .01$) which may have practical implications, if borne out in subsequent investigations. In the tactical exercise, the accuracy level for objects which interpreters reported both times they examined the imagery was significantly higher than for objects reported only once (53% vs. 34%). The result, although not supported in the case of the strategic measure, suggests that allowing interpreters to take a second look at imagery after a time lapse may prove to be useful in eliminating some erroneous identifications.

TEAM PROCEDURES

The possibility of utilizing interpreters in teams to obtain information came to mind as a matter of course. However, we did not assume that merely by working together a group of image interpreters

would automatically produce more accurate and complete identifications. We arrived at a fundamental basis for determining how the efforts of a group of interpreters could best be pooled--the concept of correlation of performance among individuals. This way of looking at team procedures appears to hold promise of real improvement in performance.

The correlation concept was applied to three different methods of obtaining intelligence information from interpreter teams. In Method I, the image interpreters worked and reported independently. The performance of two interpreters--to take the simplest case--was scored for Right identifications, Wrong identifications, and Omits (targets appearing in the imagery but not reported). Table 3 shows schematically the correlation between the two individual performances and how each element of performance--Rights, Wrongs, Omits--enters into the final product.

If there were perfect agreement between the two interpreters, every right identification made by one would also have been made by the other. The same would pertain for wrong identifications. Another way of stating this principle is that when the correlation between the performance of both interpreters is perfect, or equal to unity, there are no entries in the off-diagonal cells of Table 3. Perfect agreement, however, is obviously not what is desired when interpreter efforts are pooled. For if the interpreters agreed completely to begin with, no gain would result from pooling their efforts.

Method I could be expected to produce the maximum number of correct identifications to be obtained from the two interpreters, as shown from the computation of total scores indicated in Table 3. However, the information resulting from this method also includes all the wrong identifications made by both interpreters.

In Method II, individuals comprising the teams also worked and reported independently, but their identifications were pooled in a different way. Only identifications upon which the two interpreters agreed were considered as having been made by the team. All identifications made by only one member of the team were disregarded. The formulas for total scores under Method II (Table 3) indicate that total number of Right identifications, as well as the total number of Wrong identifications, would almost certainly be less than for Method I. Whether accuracy of interpretation with Method II is in general higher than with Method I depends upon whether image interpreters tend to agree more on the right identifications than on the wrong identifications. Since it is reasonable to suppose that interpreters will tend to agree more in reporting targets that are actually present than in reporting targets which are not present, proportionally more right identifications than wrong identifications would be expected to survive the agreement criterion established for Method II.

In Method III, interpreters worked in teams in examining the imagery, instead of working independently. Each member of a team was given complete sets of photographs, but only one answer sheet was given to a team. The team was instructed to put on the answer sheet only those identifications which both members of the team agreed upon (at least two members in the case of three-man teams). Team members could discuss the identifications freely, and go about cooperating in any way they desired.

It was expected that as a result of their discussion the interpreters would agree on some targets on which they previously disagreed. Table 4 shows the expected effect of this tendency on the elements involved in image interpreter performance correlation. We see a general movement from the off-diagonal cells to the diagonal cells and, as indicated by the heaviness of the arrows, relatively more agreement on right responses than on wrong responses. Consider, for example, targets, which one interpreter correctly identified but which the other interpreter omitted. When the interpreters compare their identifications attention will be specifically brought to these particular targets and the chances are that the correct information will be reported. We would expect considerably more agreement upon right responses for Method III than for Method II; but, owing to the fact that some right identifications will not be agreed upon, the total number of rights from the cooperative method will be less than for Method I, which pooled all right responses of the team members.

With respect to wrong responses, one would expect an even greater tendency for wrong identifications made by only one interpreter to drop out, as the other interpreter would probably be more resistant to agreeing to misidentifications. Considering that interpreters are more confident of identifications that are right than they are of their wrong identifications, there is a good chance that the interpreter making the wrong response would not argue as strongly for it as for his right responses. It is to be expected, then, that the wrongs will be fewer in number than would be obtained by Method I and that although the rights would also be fewer, the over-all accuracy will be improved, although not as much as was the case for Method II, where there was no opportunity for teammates to induce each other into agreeing to misidentifications.

How close did expected performance correspond to the empirical data? Table 5 summarizes some of the results actually obtained in the experiment. When all identifications made by two or three interpreters working independently were pooled (Method I), a much larger number of correct identifications were obtained (mean of approximately 15 as against 8 for individual interpreters.) However, by so pooling all identifications, many wrong identifications were retained. When, instead of pooling all identifications only those identifications were retained which the independent interpreters agreed upon (Method II), a marked increase in over-all accuracy

resulted. In the three performance measures used here, accuracy level rose to 33%. However, many fewer correct identifications were retained--only about four, on the average. Notice the enormous drop in number of wrong responses when only the agreed-upon identifications are considered. Apparently, interpreters do not tend to agree on wrong identifications--erroneous responses are idiosyncratic, that is, peculiar to the individual. If one interpreter calls a tonal difference on the road a vehicle, the chances are another interpreter will not agree with him, although the second interpreter may call another tonal difference a vehicle.

These data conform closely to the expected distribution of performance scores outlined earlier. As was anticipated, in team work methods, the dilemma again arises that a method which achieves high accuracy also markedly reduces the amount of accurate information extracted. If the only identifications accepted are those agreed upon by interpreters working independently, both the number of erroneous identifications and the number of targets correctly identified are markedly reduced. On the other hand, the method of pooling all reports regardless of agreement yields the most right identifications but also gives the most wrong identifications. The cooperative method, in which interpreters in a team discuss their identifications freely, appeared to offer a reasonable compromise between the extremes of the methods just outlined. Definite improvement over individual interpretation in terms of rights, wrongs, and accuracy scores was achieved by the cooperating teams (see Table 4).

We have since conducted additional studies using 2 and 3-man teams employing a variety of cooperative team work methods. Analysis of the more recent data has not been completed. However, in further analysis of our original team performance data, an interesting result was obtained when we compared individual performance of cooperating 2 and 3-man teams for equal man-hours of effort. Individuals worked alone on each measure for 1 hour or 3-man hours altogether. Their performance was compared with that of cooperating 2-man teams after the latter had worked for 30 minutes on each of the three measures and with the 3-man teams' performance after working 20 minutes on each of the measures.

Table 6 shows the results of this comparison. The 3-man teams, who only had 20 minutes to look at the photos in each measure, made about half as many correct identifications on the average as did individuals working alone. However, the 3-man teams made only about a third the errors individuals made. The 2-man teams seem to have worked most efficiently in the particular time periods compared. While they made fewer right identifications than did interpreters working alone, they averaged many fewer wrong identifications. It should be pointed out that these comparisons were made for particular time periods; comparisons for other periods of equal effort must also

be made. These data suggest, however, that when an operational interpreter group has several batches of imagery to handle, it may be more efficient to assign several interpreters to work together on each batch for a short period of time than to assign one interpreter to each batch for a longer period of time.

SUMMATION

All the techniques described above show promise of being useful in improving image interpreter performance. Plans are now being formulated to study combinations of these techniques to see whether the beneficial effects are additive. For example, experiments are currently being conducted with teams of image interpreters of varying size and composition examining imagery for a number of different time periods. Procedures will be tried out whereby individual interpreters--or teams of interpreters--view imagery once, twice, or even three times, each time assigning certitude levels using revised confidence scales. (These studies will be conducted using experienced as well as graduating interpreters.) Through these more comprehensive studies we hope to provide useful guidelines to the intelligence community on how image interpreters may be more effectively utilized.

It must be remembered, however, that technological innovations will continually affect the role of the human interpreter in the aerial reconnaissance systems of the future. We may never arrive at the one best combination of techniques for maximizing correct identifications and minimizing wrong identifications. We hope to make significant strides in this direction through continuous studies of significant aspects of interpreter performance.

For Army scientists, possibly the major interest in our findings, lies in highlighting the importance of human cognitive, perceptual, and experiential factors in system's output. The development of a visual display whose physical properties such as resolution and contrast are optimal is but one necessary step in the design of a sensor system fully responsive to military needs. The skills and techniques image interpreters use within the system must be optimized as well.

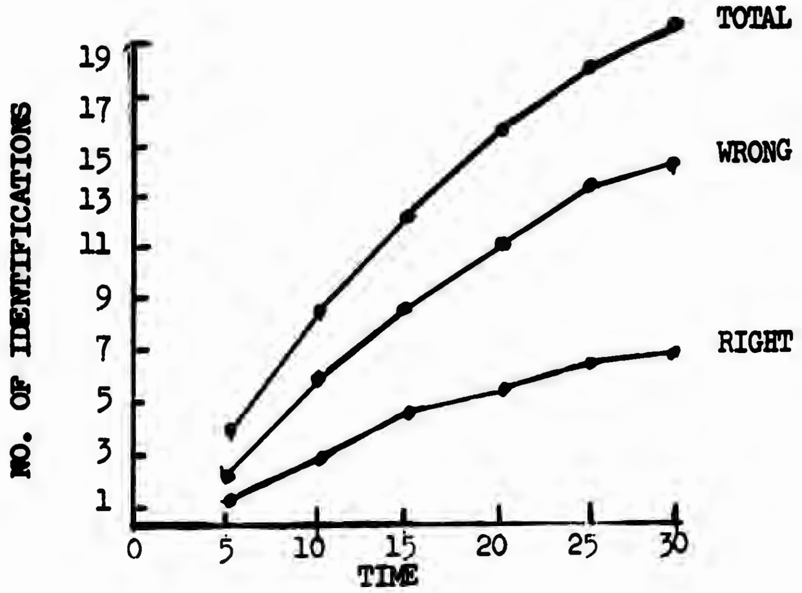


Figure 1. Representative Cumulative Curves For Right, Wrong And Total Identifications

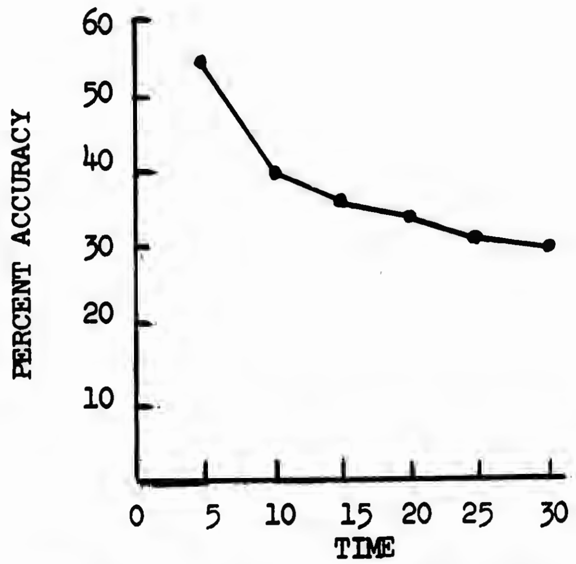


Figure 2. Representative Cumulative Accuracy Curve

Table 1

MEAN CONFIDENCE ASSIGNED TO RIGHT AND WRONG IDENTIFICATIONS ON FIVE PERFORMANCE MEASURES
(Positive = 1; Probable = 2; Possible = 3)

Performance Measures	Mean Confidence Right	Wrong
Tactical - 1	1.32	1.99*
Tactical - 2	1.54	1.86*
Tactical - 3	1.39	1.81*
Strategic - 1	1.31	1.70*
Strategic - 2	1.23	1.94*

*Significantly different from rights ($P < .01$)

Table 2

PERCENT ACCURACY OF IDENTIFICATIONS MADE WITH TWO CONFIDENCE SCALES

Confidence Scale Values	Tactical Measure 3-pt. Scale (N=58)	5-pt. Scale (N=57)	Strategic Measure 3-pt. Scale (N=58)	5-pt. Scale (N=57)
Highly Positive	--	37	--	81
Positive	32	20	71	66
Highly Probable	--	11	--	46
Probable	13	11	45	38
Possible	07	05	36	36
Over-all Accuracy	18	15	58	61

Table 3

PATTERN OF RESPONSES FOR TWO INTERPRETERS:

WORKING INDEPENDENTLY
(Methods I and II)

INTERPRETER j

INTERPRETER i		RIGHTS	OMITS	WRONGS	Σ
	R	R_{ij}	R_iO_j	R_iW_j	R_i
	O	O_iR_j	O_{ij}	O_iW_j	O_i
	W	W_iR_j	W_iO_j	W_{ij}	W_i
	Σ	R_j	O_j	W_j	

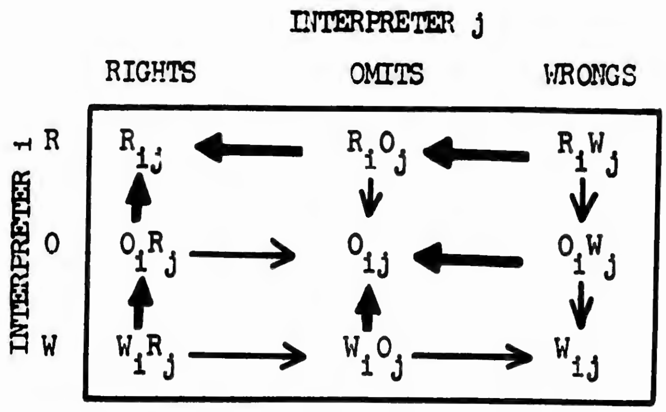
TOTAL SCORES

$R_T = R_i + R_j - R_{ij}$	(Method I)
$W_T = W_i + W_j - W_{ij}$	(Method I)
$R_T = R_{ij}$	(Method II)
$W_T = W_{ij}$	(Method II)

Table 4

PATTERN OF RESPONSES FOR TWO INTERPRETERS

WORKING COOPERATIVELY
(Method III)



Legend

R_{ij} = targets both interpreters got right

W_{ij} = targets both interpreters got wrong

O_{ij} = targets both interpreters omitted

R_iO_j = targets Interpreter i got right but Interpreter j omitted

R_iW_j = targets Interpreter i got right but Interpreter j got wrong

Other elements in matrix similarly defined

Table 5

TOTAL MEAN PERFORMANCE SCORES FOR TEAMS WORKING UNDER DIFFERENT METHODS AND FOR INDIVIDUALS WORKING ALONE

Team Methods*	Right	Wrong	% Accuracy
Independent: All responses (N = 15)	15.1	112.9	14
Independent: Agreed responses (N =15)	3.9	7.1	33
Cooperative (N = 30)	10.1	40.6	22
Individuals Alone (N = 36)	8.1	50.8	16

*Mean Right, Wrong, and Accuracy scores between team methods are significantly different ($P < .01$).

Table 6

MEAN PERFORMANCE SCORES FOR INDIVIDUALS AND TEAMS WORKING A TOTAL OF 3-MAN HOURS ON 3 PERFORMANCE MEASURES

	Time (each measure)	Right	Wrong	% Accuracy
Cooperating 3-Man Teams (N = 12)	20 Min.	4.2	17.9	22
Cooperating 2-Man Teams (N = 9)	30 Min.	7.6	27.6	22
Individuals Working Alone (N = 36)	60 Min.	8.1	50.8	16
Significance of Mean Differences		$P < .05$	$P < .01$	n.s.

THE ERROR FIELD ASSOCIATED WITH INSTRUMENTATION FOR POSITION DETERMINATION

TH.-W. SCHMIDT
U. S. ARMY RESEARCH OFFICE (D)
DURHAM, NORTH CAROLINA

Problems of position determination in astronomy and geodesy originated the establishment of the method of least squares by Legendre and Gauss at the beginning of the 19th century. The concept of the ellipse of error (or ellipsoid of error in the three dimensional case) as the most complete way to describe the precision of a position determination was subsequently first introduced by the French scientist Bravais with his paper, the year 1846, "A mathematical analysis of the probabilities of error in the location of a point" (1). This was followed by contributions from Helmert (2) and Schols (3). Those three early publications covered all essential aspects of this concept. However, the ellipse of error, although applied sometimes in the field of geodesy or ballistics and described in textbooks on data adjustment, never gained many applications in the past. More than a hundred years had to elapse until Shapiro (4) used in his book "The prediction of ballistic missile trajectories from radar observations" the ellipse (or ellipsoid) of error as the only estimator for the precision of the prediction.

All the numerous publications from 1846 to 1957 on error ellipses, however, have one surprising fact in common: they consider only the error of a single point. Test planning and systems evaluations today require, however, information on the coverage of the whole test area. Point by point calculation and subsequent interpolation are out of question, as even modern computing equipment cannot handle the amount of numbers involved for any proving ground. A new approach was needed and the concept of the error field as the entirety of all points covered by the measuring system was proposed in 1958 by the author (6).

As the differential equation for the error field is obtained directly from the normal equations for the data adjustment, it is hard to believe that this interpretation was not proposed long before in one of the numerous publications on error analysis, based on

the method of least squares. However, no article has been found so far, which advocates the concept of an error field. The author therefore believes, with all reservations, that the idea of an error field may be one new contribution in this area.

This paper will demonstrate the field concept in the two dimensional case, where the error field is described by a quadratic differential equation. Its solution results in a system of orthogonal curves and the axes of the ellipses of error are at each point tangents and normals to those curves. As an illustration, examples of error fields for several different types of measurements for point location in a plane will be presented.

It has been known since the earliest publications on the subject, that the semi-axes of the ellipse of error are obtained, save for a scale factor, from the eigenvalues of the coefficient matrix of the normal equations, derived by the method of least squares for the adjustment of data with the measuring system investigated. The scale factor is the error in an observation of unit weight; it may be estimated independently, as in the theoretical considerations here, or, as in the practice of data reduction, the redundancy of the observations will allow its estimation. (5)

Writing the coefficient matrix of the two normal equations in the form

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}$$

(associating as usual the first column with the x axis, the second column with the y axis of an arbitrarily oriented, cartesian coordinate system) the size of the semi-axes (a,b) of the ellipse of error follows from the two, always real and positive roots of the quadratic

$$(1) \quad \lambda^2 - (a_{11} + a_{22})\,\lambda + (a_{11}\,a_{22} - a_{12}{}^2) = 0,$$

so that $(\frac{a}{\sigma_1})^2 = \frac{1}{\lambda_1}$ and $(\frac{b}{\sigma_1})^2 = \frac{1}{\lambda_2}$, where σ_1 is the error of an observation of unit weight. It has been proved that a and b are invariant quantities and independent of the coordinate system used.

The orientation of the axes of the ellipse of error is obtained from

$$(2) \quad \cot 2\,\psi = \frac{a_{11}-a_{22}}{2a_{12}}$$

where ψ and $\psi + \frac{\pi}{2}$ are the angles between the semi-axes (a,b) and the

x-axis of the coordinate system used. It should be mentioned, that the orientation of the ellipse of error is invariant with respect to the coordinate system and depends only on the geometry between the point investigated and the measuring stations. (Corresponding formulas for the three dimensional case may be found in reference 4.)

The coefficients a_{ik} in the normal equations are functions of the coordinates (x,y) of the point considered and contain the station coordinates as constants. Equation (2) can therefore be written as

$$(3) \quad \cot 2\psi = g(x,y),$$

and by means of well known trigonometric formulas the quadratic equation for the angle ψ results:

$$(4) \quad \tan^2\psi + 2\, g(x,y) \tan\psi - 1 = 0.$$

Equation (4) establishes in each point of the x-y plane the directions of the semi-axes of the ellipses of errors. Drawing those directions for the entirety of points establishes a directional field, which in the limit leads to the error field itself. Let $\tan\psi = y'$; then, from (4)

$$(5) \quad y'^2 + 2\, g(x,y)\, y' - 1 = 0 \qquad \text{or}$$

$$(6) \quad y' = -\, g(x,y) \pm \sqrt{1 + g^2(x,y)}$$

is the differential equation for two families of curves. These two families will constitute the orthogonal system of the error field. At any point the tangents and normals to those two curves, passing through the point will coincide with the directions of the axes of the ellipse of error at this point.

By virtue of the normal equations and the least square approach for data adjustment there exists always such a system of orthogonal curves, no matter how many stations are employed in the measuring system or how the stations are located. The orthogonal system of the error field characterizes completely the measuring system and may be regarded as its "natural" coordinate system. As shown later, systems, measuring in their natural coordinates, have certain advantages.

The simplicity by which the differential equation of the error field is obtained from the equation for the double angle of rotation of the ellipse of error in a point, made it difficult to imagine that this field concept had not been proposed long before. However, the numerous papers on the method of least squares and its applications make a complete check very difficult. The author states therefore only with reservations, that in his belief this concept of the error field associated with instrumentation for position determination may be a belated new contribution in the area of error analysis, originated more than a century ago.

As illustrations of the concept of error fields in the plane four examples of different measuring systems will be presented. Three of the examples have relation to systems, utilized in practice; the fourth example shall serve as a demonstration, that the mean point error, i.e. the invariant quantity of the sum of the squares of the errors in two perpendicular directions, which equals the sum of the squares of the semi-axes of the ellipse of error, is not a unique quantity of a measuring system.

The oldest method of determining the location of a point in a plane consists in triangulation, or the measurement of two angles from two stations, located at the ends of a base line of known length. Without loss of generality a cartesian coordinate system may be centered at the midpoint of the base line, the x axis in the direction of the base line, and the stations located at (-1,0) and (1,0) respectively. A point (x,y) in the plan is then observed from the stations under the angles

$$\alpha = \text{arc tan } y/(x + 1) \text{ and } \beta = \text{arc tan } y/(x - 1)$$

The standard procedures of the method of least squares for non-linear equations of observations (5) utilizing Taylor's expansion and retaining linear terms only, yield the normal equations and subsequently the differential equation for the error field (5)

$$y'^2 + 2\, g(x,y)y' - 1 = 0,$$

$$\text{where } g(x,y) = \frac{(x^2-y^2 + 1)[(x^2 + y^2 + 1)^2 + 4x^2] - 8x^2(x^2 + y^2 + 1)}{2xy[(x^2 + y^2)^2 - 2(x^2 - y^2) - 3]}$$

No solution in closed form has been found for this differential equation and the author is indebted to Dr. T. Gallie, Director of the Computing Laboratory at Duke University, for the numerical solution, performed in his laboratory.

Figure 1 is the graphical representation of the error field for this old problem of triangulation. As further illustrations, some ellipses of error in proper orientation and proper relative size are shown. The size of the error ellipses is of course exaggerated and has no relation to the length of the base line between the two stations.

The error field is symmetrical with respect to the x- and y-axes, and it resembles to a certain extent the known orthogonal system of Cassinian ovals and hyperbolas. The error field can indeed in the far distance be approximated by the system of Cassinian ovals, as all terms of the sixth order of x and y or xy in g(x,y) are the same for both of the orthogonal systems. In the neighborhood of the base, however, the systems differ, and instead of the separating lemniscate in the Cassinian system here exist two separatrices, enclosing the points (-1,0) and (1,0) each in a loop, but open in opposite directions. The allotted time and space for this paper excludes any more detailed discussion of the error field. It may only be noted for further reference, that error ellipses of equal size exist only in

pairs or quadruples according to the symmetry conditions; here exist no loci for such equal error ellipses and it is therefore not possible to traverse this error field on a path which could be measured with constant precision.

The second example presents the complementary case for bipolar coordinates, consisting in the measurement of the two distances r_1 and r_2 from two stations, located at the end points of a baseline as in the first case. This example was treated by the author in another paper (6) in detail. The derivation of the differential equation will therefore be omitted, and it will only be noted that in this case the error field can be established by geometrical considerations without setting up its differential equation. Figure 2 shows the result: the error field in this case of measuring distances with constant error of measuring consist of the classical orthogonal system of confocal ellipses and hyperbolas. As before, some included ellipses of error illustrate relative size and orientation. It may be mentioned that here exist loci for error ellipses of equal size, as stated in (6) and it is therefore possible to have paths through this error field for measurement with equal precision. Those loci are circles through the endpoints of the base line.

The two presented cases of measurement in bipolar coordinates have a common feature: no measurement is possible on the baseline itself or on its extension. This is obvious from geometrical consideration and due to the linearization for the normal equations the region near this line should be excluded as a taboo-region, as is done in practical applications.

The third system presented follows an investigation of Jordan (7) with slight modification of his measuring concept. Figure 3 shows the error field for point location by means of cartesian coordinates: measuring the distances x_i and y_i on the axes and determining the point (x_i, y_i) by the intersection of the two perpendiculars to the axes, erected in x_i and y_i.

The error field consists of parallels to the two axes and the size and orientation of error ellipses are indicated in the figure. The quantity of interest in this case is however the mean point error, defined as $\sigma^2_m = \sigma^2_x + \sigma^2_y \equiv a^2 + b^2$

This quantity is easily obtainable in every case from the normal equation and by virtue of its invariance of the coordinate system, it is today nearly exclusively used for a description of the error behavior of a measuring system. The equation in the figure derives the loci for constant mean point error as circles with common midpoint at the origin.

The last system shown is the error field for point location by means of polar coordinates. The shape of this error field follows immediately from the measuring principle, measuring radial distances

and the angles of the radius vector with respect to some reference direction, both measurements performed with constant precision. The left half of figure 4 shows the error field, given by concentric circles and the radii vectores. The loci for ellipses of equal size and therefore also of constant mean point error are also concentric circles with the origin at the center, they are therefore the same circles as in the case of the cartesian coordinate system, despite the completely different error fields in the two cases.

This fact shows clearly, that the commonly used mean point error is alone not sufficient for a description of the error behavior of a position measuring system. The discussed cartesian coordinate system was an easy means to give this example.

The discussion on possible applications of the error field in the remaining part of this paper will be restricted to the three measuring systems utilized in practice: polar coordinates for a single station instrumentation, or for two station systems the measurement of two angles or two distances. All three systems approach, at distances very far from the origin, the same error field configuration of concentric circles and radii vectores; however for the angular measuring system the major axis of the error ellipse is oriented in the direction of the radius vector, for the two distance measuring systems perpendicular to the radius vector. (This behavior has its counterpart in the three dimensional case, to be treated in another paper, where it will be shown, that the general shape of the ellipsoid of error depends on whether closed or open quadratic surfaces intersect in the point under consideration.)

The polar coordinate system has in comparison with the two other systems a distinguishing advantage: it measures in its "natural" coordinates, meaning in the orthogonal coordinates of its error field. This means that any difference in the precision of the angular or the distance measurement may change only the size and the shape of the error ellipse at the point measured, but not its orientation. That error field as such is stable and will not change its configuration.

This is not the case with the two two-station systems discussed. Rewriting their normal equations with different weights for each station results in different error fields, dependent on the weight assumptions. In the case of distance measurement however it is possible to stabilize the error field by changing the method of measuring. Instead of measuring separately r_1 and r_2, the sum r_1+r_2 and the difference $r_1 - r_2$ may be measured. From geometrical consideration the error field of confocal ellipses and hyperbolas remains the same. However the measurements are now performed in "natural coordinates", and changes in the measuring performance of those two quantities will not change the error field configuration. Electronic interferometer systems are based on the measurement of the difference r_1-r_2; however for convenience r_1+r_2 is not measured but the radial distance from the midpoint of the baseline. This, unfortunately,

leads to an error field that is not stable.

The final purpose of the error field for an instrumentation system is of course to be an aid in predicting errors for trajectory determination. Here usually statements of errors in a given direction, as tangential or normal to the trajectory, are required.

The error at one point in any given direction having the direction cosines l and m respectively to the semi-axes of the ellipse of error at this point is known to be

$$\sigma^2_d = a^2l^2 + b^2m^2.$$

Maximum precision is of course obtained when the direction coincides with the minor axis of the error ellipse, i.e. when the relative orientation of trajectory and instrumentation is such, that the important region of the trajectory is essentially coincident with the field line for the minor axis of the error ellipses. It is sometimes possible to make such an arrangement, but it may not always be the most desirable orientation. First, this orientation, which gives minimum error in one direction, results in maximum error in the direction perpendicular to the first. Requiring maximum precision in the tangential direction automatically results in the lowest precision for the normal direction. Secondly, selecting flight paths in the direction of the minor axis has another drawback: any small change from the exact direction of the minor axis leads to a rapid increase of the error. The direction of the minor axis is indeed the direction in which this change has its maximum. Such changes of direction may occur either by a slight deviation of the missile trajectory from its planned course, or they may be the result of differences in performance of the instrumentation in cases of error fields that are not stable.

This leads to the question: are there flight paths possible traversing the error field, which are the least sensitive to such variations? It can be proved that such trajectories always exist, and that they always bisect the angles between the axes of the ellipses of error. They have therefore also the property, that their errors in tangential and normal directions are equal at each point; however they differ from point to point.

From the conditions for isogonal trajectories the differential equation (5) of the error field yields immediately the differential equation for these "insensitive" trajectories:

$$(7) \quad y'^2 - \frac{2}{g(x,y)}\, y' - 1 = 0$$

with $g(x,y)$ as before. This differential equation can always be solved by numerical integration and defines another orthogonal system, rotated in each point by $\frac{\pi}{4}$ with respect to the error field.

For the polar coordinate systems as an example, those "insensitive" trajectories can be easily established: The axis of the error ellipse must always be intersected at $\frac{\pi}{4}$, but one of the axis

is the radius vector, to be intersected under a constant angle. The required trajectories consist therefore of two families of logarithmic spirals, which are orthogonal to each other. Arranging the relative position of a radar to the missile flight path in such a way that the important section of the flight path is approximated by one of the "insensitive trajectories", i.e. logarithmic spirals, will result in a system with high repeatability in its daily performance.

One point remains to be discussed: how to include the influence of refraction on the error field. Point by point correction, as needed for data reduction, would of course spoil the approach presented here of describing the whole error field by differential equations. To include the general influence of refraction in the original equations and to proceed as before is theoretically possible, but would lead to very complicated formulas. Another approach may therefore be preferable: The transformation of the error field, derived without refraction, by suitable mapping into the error field which includes the refractive effect. This way was suggested by the author (11) and the mapping concept as a means to deal with problems of wave propagation in inhomogeneous media is also proposed by Ivanov(12). The difficulty lies of course in the need for the appropriate mapping function, which is to date not known for the general case.

The basic idea can however be demonstrated for a special case, consisting in the approximation of the true ray path by employing its radius of curvature at the station, as used for a long time in electronics. For a single station system, such as a radar station, the problem can be reduced by virtue of rotational symmetry to a two dimensional treatment in a vertical plane, containing the station and the center of the earth. From the known properties of refractive curves for a spherical earth and spherically stratified refractive index (10) it follows that the endpoints of all radii of curvature lie on a line, perpendicular to the radius of the earth through the station. This condition corresponds to the well known transformation $w=C \ln (z-a)/(z+a)$. The right side of figure 4 shows the effect of mapping by this function: the previous error field of radii vectores and concentric circles transforms through the influence of refraction into a new error field consisting of orthogonal intersecting circles. This new error field may now represent very closely the situation in a real atmosphere, as long as the approximation of the ray path by employing its radius of curvature is valid.

Summarizing in conclusion: This paper presents the idea of describing the errors in position determination by the error field associated with the instrumentation. It shows that the differential equations for this field can be derived from the normal equations for data adjustment, and proposes finally, that the influence of refraction may be included by suitable mapping of the error field, obtained for the case without refraction. The examples are given for position location in a plane, and are also valid for position determination in space, if the system has rotational symmetry. The extension of the

field concept to three dimension will be the object of further studies.

The author has had many helpful discussions with Dr. A. S. Galbraith while this work was in progress.

REFERENCES:

1) A. BRAVAIS, "Analyse Mathématique sur les Probabilitiés des Erreurs de Situation d'un Point." Mémoires Presentes par Divers Savants a l'Académie Royale des Sciences de l'Institut de France. Vol. 9, 1846, pp. 255-332.

1a) English Translation: "A Mathematical Analysis of the Probabilities of Error in the Location of a Point" by A. Bravais, published February 1958, White Sands Missile Range, N.M. (Translated by Th-W. Schmidt, J.G. Wozencraft and M.M. Hayes)

2) F. R. HELMERT: "Studien ueber rationelle Vermessungen im Gebiete der hoeheren Geodaesie." Zeitschrift fuer Mathematik und Physik. Vol. 13, 1868, pp. 73-120; 163-186.

3) CH. M. SCHOLS, "Théorie des Erreurs dans le Plan et dans l'Espace" Annales de l'École Polytechnique de Delft, Vol. 3, 1886, pp. 123-179 (Presented to the Academy of Sciences of Amsterdam November 1874)

3a) English Translation: Theory of Errors in a Plane and in Space" by Ch.M.Schols, published April 1958, White Sands Missile Range, N.M. (Translated by Th-W.Schmidt, J.G. Wozencraft and M.M. Hayes.)

4) I.I. SHAPIRO, "The Prediction of Ballistic Missile Trajectories from Radar Observations" February 1957,(McGraw-Hill Book Co.) New York, N.Y. pg. 85 ff and 108 ff.

5) E. WITTHAKER and G. ROBINSON, "The Calculus of Observations" 1946, (Blackie and Son Ltd) London, pg. 214.

6) TH-W. SCHMIDT and MARTHA M. HAYES, "Methods of Error Analysis in Two Dimensions. June 1958. Special Report OOR Project I-170-M, White Sands Missile Range, N.M.

7) E.E. PERRET, "Error Analysis for Determination of Target Position and Velocity from two or more Observers. June 1960. WADD Technical Note 60-70, Wright-Patterson Air Force Base, Ohio.

8) W. JORDAN, "Ueber die Genauigkeit einfacher geodaetischer Operationen" Zeitschrift fuer Mathematik und Physik, Vol. 16, 1871, pp. 397-427.

9) O. EGGERT, "Ueber die guenstigste Punktlage beim Einschneiden". Zeitschrift fuer Mathematik und Physik, Vol. 49, 1903, pp.145-168.

10) J. A. GRUNERT, "Ueber die atmosphaerische, vorzueglich die Terrestrische Refraction, und ueber Refractionscurven im Allgemeinen." Archiv der Mathematik und Physik, Vol. 10, 1847, pp. 1-41.

11) TH-W. SCHMIDT: "Ueber Genauigkeit und Fehlertheorie bei Schnittpeilungen in der Ebene und im Raume." Sonderbuecherei der Funkortung: Genauigkeit und Reichweite von Funkortungsverfahren. Vortraege und Diskussionen der internationalen Diskussionssitzung in Muenchen am 25. und 26. April 1960. pp. 106-117. (No. 2035, Ausschuss fuer Funkortung, Duesseldorf)

12) V. I. IVANOV, "Application of Conformal Mapping to the Simplest Problems of Wave Propagation in Inhomogeneous Media" Zh.Vychisl.Mate.Mate.Fiz. Vol. 1, 1961, pp. 246-254.

(English Translation: ASTIA AD-262 425).

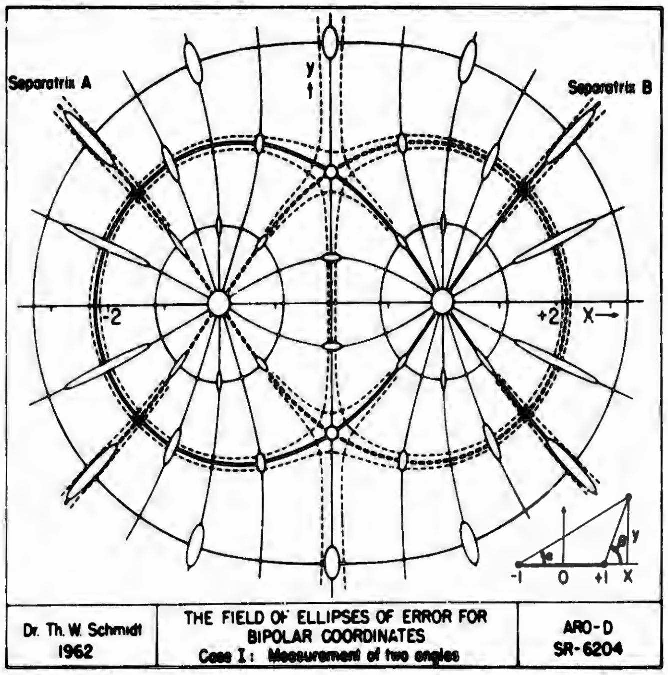


Figure 1

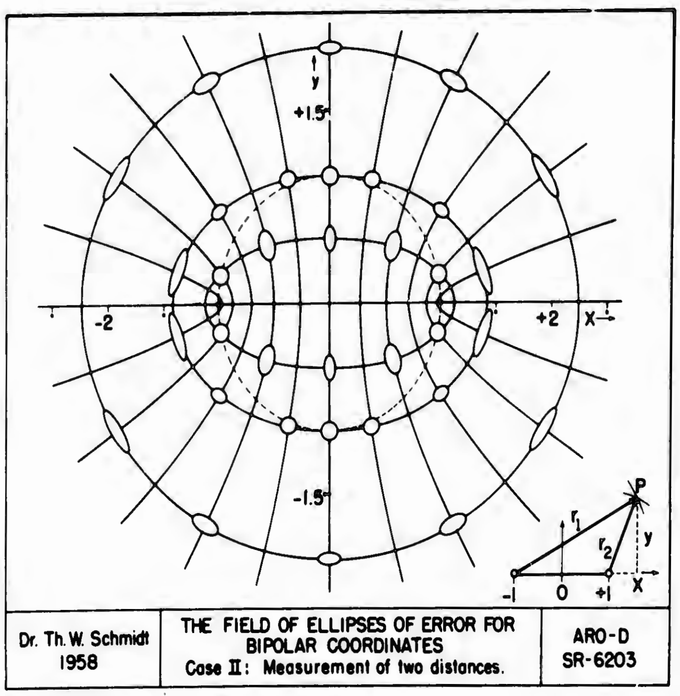
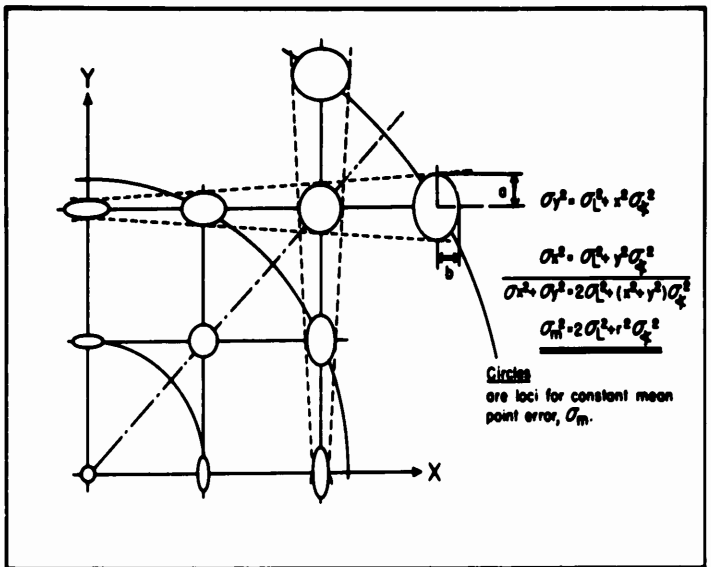
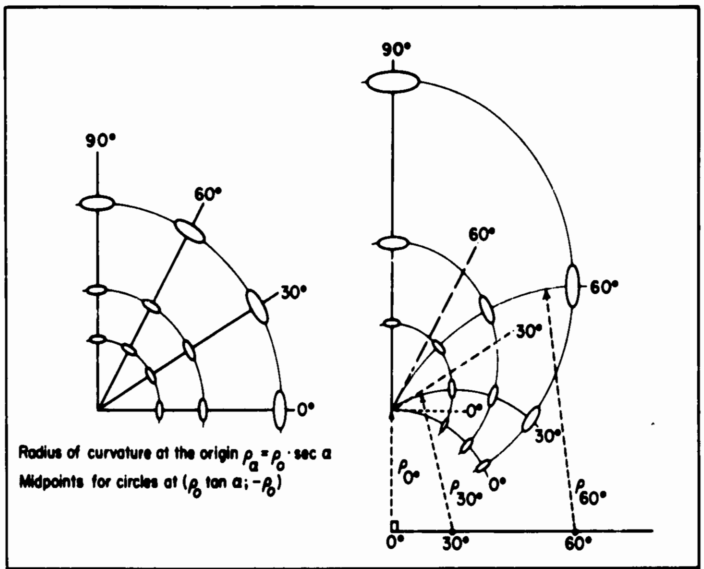


Figure 2



Dr. Th. W. Schmidt 1960	THE FIELD OF ELLIPSES OF ERROR FOR CARTESIAN COORDINATES Perpendiculars on two orthogonal axes	ARO-D SR-6201

Figure 3



Dr. Th. W. Schmidt i960	THE FIELD OF ELLIPSES OF ERROR FOR POLAR COORDINATES AND ITS RANSFORMATION BY MAPPING	ARO-D SR-6202

Figure 4

A MODEL FOR EVALUATING SURFACE-TO-SURFACE FIREPOWER

OTIS S. SPEARS
OFFICE OF COMBAT DEVELOPMENT AND DOCTRINE, USAAMS
FORT SILL, OKLAHOMA

1. INTRODUCTION.

a. To the extent that warfare becomes an instrument of national security, methodical selection and development of weapons must be engendered. Thus, there must be a sound basis for weapons design and analysis. Ineffective weapons, however skillfully employed, can result only in disaster on the battlefield. The purpose of this discussion is to outline a methodology, a perspective for the determination of weapons effectiveness. While certain general principles mentioned may pertain to all weapons, the immediate consideration is for surface-to-surface tactical weapons.

b. Basic to the entire discussion is the assumption that tactical weapons must be designed and measured with respect to overall effectiveness on the battlefield, not just from the standpoint of some particular desirable or undesirable weapon characteristic. More specifically, weapons are examined with respect to total firepower potential, measured and summed over different enemy tactical configurations. Each weapon is considered as being potentially a single weapon, a portion of a tactical unit; and an element of a whole force structure. The particular factors involved are range of the weapon(s), nature and size of enemy targets, enemy tactical dispositions, rate of fire, weapon accuracy, and the potential destructiveness of a single projectile. Closely allied considerations are weapon reliability and mobility. These factors are interwoven in such a manner as to yield an overall measure of tactical effectiveness.

2. GENERAL DISCUSSION.

a. Historically, weapons development has proceeded along lines which have been usually

intuitively felt to lead toward more effective fire. Actually, these lines were largely dictated by specific developments in other areas. Prior to the coming of the nuclear weapons, however, there was no specific attempt to evaluate the effectiveness of a weapon or group of weapons.

b. With the advent of nuclear warheads, a procedure for effects analysis was evolved. This procedure was derived by the Sandia Corporation at Albuquerque, New Mexico. The Command and General Staff College adapted the theory and methodology to instructional needs in special weapons courses. These target analysis procedures are now in general use throughout the Army (Ref 1). Only a limited number of factors are considered: nuclear yield, weapon accuracy, target size, and nature of the target (hard, medium, or soft).

c. During the last few years, weapons research and evaluation have advanced considerably. It has become increasingly evident that a more thorough methodology must be used, especially as guidance for research and development efforts. Here, the analysis must include consideration of as many tactically significant factors as possible. Perhaps no single analytical model can ever encompass the formidable array of tactical elements which invariably confronts the commander in any actual battle situation. Nevertheless, the complexities of modern weapons imply a fundamental and methodical attack on the problem of efficient weapons development. From drawing board to service test, the entire weapons development process will be futile unless it leads to sufficient firepower potential to cope with the enemy threat. Therefore, the assumption here is (indeed must be) that it is possible to develop an overall measure of tactical effectiveness for a weapon or an entire series of weapons, including such important factors as accuracy, range of the weapon, the target complex, effectiveness of an individual projectile, rates of fire, target location errors, and emplacement distances relative to enemy dispositions. Closely allied considerations are weapon reliability and mobility. The purpose of this discussion is to outline one possible approach for effecting such a measure. First, it is well to delineate certain qualitative concepts.

d. Each weapon or delivery unit (such as a battery of field artillery) is limited to some particular area within which it can fire. While this area need not be confined to any given shape, it has often been fan-shaped in the past, with the pivot at the center of the position area of the delivery system. These regions, whatever their shapes, will hereafter be called "capabilities regions." Assume that the locations of all

weapons and enemy dispositions are correctly drawn to scale on a war game map. Then, the portion of the enemy area accessible to each weapon unit is outlined simply by drawing the configuration of the capabilities region. Here it is assumed that the entire friendly and enemy complex is drawn to the same scale. The situation is represented schematically by figure 1. The entire capabilities region of the indicated field artillery battalion is shaded.

e. This procedure is continued until all accessible areas have been outlined over enemy territory for each weapon system contemplated. Within a given sector of the battlefield, these capabilities regions may thus be depicted for all weapons within an entire force structure. On the other hand, it may be necessary to materialize only those regions corresponding to weapons most tactically appropriate for the given regions. In the first instance, the total firepower potential for the entire force structure will be measured. In the second instance, only the potential of those weapons deemed appropriate for the region will be measured. Both measures have merit, for it is not always possible to determine which weapons will be used within a given battle area before the engagement is in progress. Moreover, one task of the weapons analyst is to investigate the advantages and limitations of different weapons arrangements.

f. The several capabilities regions will usually generate different kinds of overlap areas with rather irregular geometrical shapes. Figure 2 is a schematic representation of this notion. Each kind of overlap zone is classified by some simple code number. For example, the first digit might represent the number of 105mm howitzer battalions which can reach the area. The second digit might represent the number of 155mm howitzer battalions which can reach the overlap zone. Some other digit would represent the number of, say, division general support guided missiles which can reach the area.

g. Thus, the battlefield is covered with different types of capabilities regions and overlapping areas thereof. Next, each such subarea of the battlefield is examined for enemy installations. All targets may be listed, but it is usually sufficient to select only one type of target within each subarea (overlap zone category). This target is usually the one with the highest tactical priority, or the most difficult one. The size and nature of the target are then listed (e.g., a platoon of enemy infantry in armored personnel carriers within a circle having a radius of 200 meters).

h. The next step is to compute the measure of damage to a typical target (or to each target) within

each subarea or overlap zone (an overlap zone includes any subarea which can be reached by only one weapon). The damage measure is usually one of two types; first, the expected, or average, damage level; and secondly, the fractional damage which can be realized with some specified probability. In either case, the damage level is determined by the size and nature of the target, the number of projectiles fired, the accuracy of the weapon, and the damage potential of each projectile, as well as by the accuracy of target location and weapon position location. For example, when both weapon and target are accurately located, the accuracy of the weapon then determines the percentage of all projectiles fired that can be expected to fall within the confines of a given target. The net result of this entire step is the association of a damage level with each specified target throughout the battlefield. This calculation can be effected for as many points as are desired; but, as already noted, it is usually advisable to select a set of typical, or representative, targets. Here, a "point" ordinarily designates the center of a given target. Figure 3 represents typical results for several points, or target centers, within enemy territory. Even though the generation of each damage-level number is a laborious process, this information must be evolved for all subareas of the battlefield. Specific procedures for individual targets are set forth in references 1 and 2.

i. Next, contours of constant-damage potential are drawn. This is an approximation process similar to that used by weather forecasters to draw isothermal lines on a synoptic weather map. The results are also analogous to contour lines on a map, showing loci of constant elevations. The lines can be produced as accurately as availability of skilled personnel and computational facilities permit. The actual computation will be greatly facilitated by the use of digital computers.

j. It is instructive to examine some actual examples. Figure 4 shows an actual deployment of division artillery used during a recent study. Some of the specific designations have been removed to eliminate the necessity of classifying the material. Figure 5 is a recent actual example of capabilities regions and overlap zones, not derived from the deployment shown on figure 4. Only sequential zone numbers are given, rather than the code numbers mentioned previously. Figure 6 shows a recent set of damage contours. Neither the weapons deployment nor the target complex used is shown. Figure 7 shows the rather significant results obtained when certain changes were made among the weapons, such as changes in ammunition types and weapons locations.

k. Finally, these calculated capabilities are summed over the entire battlefield. Essentially, this is a summation of weighted subareas, each subarea being multiplied by an appropriate weighting factor. This weighting factor may simply be the damage-level numbers previously described, or it may include target priority. For example, targets in the rear areas may not be so important as those near the front lines. The result of the summation is the overall measure mentioned at the beginning of the discussion.

3. THE FUNDAMENTAL EQUATIONS.

a. It is well to recall the basic factors for weapons analysis already mentioned. They are as follows:

(1) Accuracy.
(2) Lethality.
(3) Rate of fire.
(4) Weight.
(5) Range.
(6) Mobility.
(7) Reliability.
(8) Size and nature of target.

b. The dispositions of friendly and enemy units, together with these factors for weapons analyses, determine the damage-level potentiality against any enemy target. The measure of this potentiality is <u>fractional damage</u> or fraction of casualties--fraction of casualties, if the target is personnel; fractional damage, if the target is other than personnel. For any given weapons plan and any specified target, let this destructive potential be denoted by $G(x,y)$, where (x,y) represents the location of some designated point within the target, usually the center. Then, the value of $G(x,y)$, associated with the location (x,y), is a fraction of damage, or a fraction of casualties, which the weapons can inflict on the given target when they are aimed at the point (x,y). Then, for any point over the entire battlefield, there is some value of $G(x,y)$ such that,

$$0 \leqq G(x,y) \leqq 1. \tag{1}$$

The sum total of this potential over a given battle area is the overall measure of weapons potential which was qualitatively outlined in the first portion of the discussion. Let M represent this total measure. Then, by definition,

$$M = \iint_{(R)} G(x,y)\, dx\, dy. \tag{2}$$

Here, the symbol $\int\int_{(R)}$ indicates that the integration is to be performed over the battle region (R). The dimentions of M are those of area; hence, results are exhibited as, say, square meters.

c. It was noted in the first portion of the discussion that the value of G(x,y) may be modified by the target priority in the determination of the overall measure M. This priority may be constant over the battlefield; it may be arbitrarily designated as a function of range; or it may be a function of range, nature of target, and other factors. Let this priority factor be called f(T), where,

$$0 \leqq f(T) \leqq 1 \tag{3}$$

Then, equation (2) is written as follows:

$$M = \int\int_{(R)} G(x,y)\, f(T)\, dx\, dy \tag{4}$$

It is understood that the value of f(T) is that which is associated with the target centered at (x,y) or otherwise designated by (x,y). It should be recalled, also, that the value of G(x,y) may be computed as the <u>expected damage level</u> for the target associated with the point (x,y); or this value may be the damage level associated with some designated probability. In the latter case, the probability referred to is that of achieving <u>at least</u> the damage level G(x,y). These two different approaches will yield different values of M. It appears that tacticians generally prefer the latter approach, with a high associated probability.

d. The measure of weapons accuracy is the probable error, circular or linear. These are standard measures of dispersion about the mean. They are so well known that further elaboration is not necessary here. For surface-to-surface weapons, however, it should be noted that the term "range probable error" designates the probable error in the direction of fire; "deflection probable error" is the probable error along the axis perpendicular to the direction of fire. The accuracy patterns of individual weapons are usually assumed to be sufficiently described by the normal, or "Gaussian," probability density function. The composite distribution function for several weapons aimed at the same target point may, however, not be "normal." In that case,

it is often rather difficult to write an analytical function adequately describing the distribution. Simulation techniques and numerical-analysis procedures can be used.

e. The measure of effectiveness for the individual projectile is the damage radius (R_D) for nuclear weapons, the lethal area (A_L) for high explosive projectiles. The nuclear damage radius is the radius within which at least some specified damage level is expected to occur. The associated probability at the outer end of the damage radius is 0.5. For high explosive projectiles, the lethal area is the probability-of-casualty integral computed over the damage pattern of the projectile, with respect to a given target situation (such as personnel in uncovered foxholes). Damage radii for different nuclear yields can be found in many documents on nuclear weapons, including Reference 2. Since the concept of lethal area for nonnuclear projectiles is perhaps less familiar than that of damage radii for nuclear weapons, a further remark about lethal area is in order. This extremely useful concept was derived and promulgated by representatives of the U. S. Army Ordnance. Suppose a high explosive projectile detonates at (or above) any point within the target area. Let this point be the origin of coordinates (0,0). The probability of a casualty at any point (x,y) is determined by the fragmentation characteristics of the projectile, as well as by the disposition and nature of individual targets. Let this probability be denoted by g(x,y). Here, the "probability of a casualty at any point (x,y)" indicates the probability that an individual target element with a finite presented area, with its center at (x,y), will be a casualty. Let the symbol for lethal area be denoted by A_L. Then, by definition,

$$A_L = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)\, dx\, dy \qquad (5)$$

Actual lethal area data are usually supplied by the Army Ordnance.

f. Actually, the overall measure M, given by equation (4), is usually approximated by a numerical analysis procedure of some sort. One such procedure has already been generally indicated by the figures previously shown. Thus, types of overlap zones on the battlefield are developed, and perhaps one type of target is assigned to each such zone. For n overlap zones (or any other desirable scheme of subdividing the battlefield), equation (4) is then written as follows:

$$M = \sum_{i=1}^{n} G_i(x,y) f_i(T) \tag{6}$$

where, $G_i(x,y)$ denotes the value of $G(x,y)$ for the i-th subdivision of the battlefield. Similarly, $f_i(T)$ represents the corresponding value of $f(T)$, for the same i-th subdivision.

g. Figure 8 shows plots of the measure M for several different tactical deployments. Certainly, the lowest value represents a less desirable tactical deployment than any other case shown.

h. Finally, values of M, or even G(x,y) can be exhibited for various intervals through an enemy battle array. The result is a broken-line graph of overall performance extending through the enemy area. Here, it is necessary to choose arbitrary increments of range throughout the battlefield, beginning usually with the forward edge of the battle area. In this case, values of M or G(x,y) are portrayed along the vertical axis, while the horizontal axis shows range, usually in kilometers, through the battle area. Figure 9 illustrates this type of portrayal for values of G(x,y). This kind of delineation is perhaps best adapted to situations in which relative comparisons are desired among individual delivery units, especially when values of G(x,y) are exhibited on the vertical axis. Actually, figure 9 is a recent example of such comparisons among certain surface-to-surface missile systems.

i. In this short discussion, concepts and measures have been only briefly and generally outlined. For actual implementation, it is necessary to extend and further develop this study. For example, the allocatior of specific weapons to different tactical missions and different portions of the battlefield has not been sufficiently discussed. In this regard, an important consideration is the total value of the quantity M which can be developed within a given time interval by some particular allocation of weapons firing simultaneously over different portions of the enemy area. Another interesting quantity is the rate of change of M with respect to time (dM/dt) at stipulated times after firing begins. Another tactically significant further consideration involving mobility relates to the change of the quantity M when a given portion of the force is ordered to move, and thus becomes unable to fire until the move is completed.

4. CONCLUSION.

a. A model for evaluating surface-to-surface firepower has been generally outlined. The results are measures of total firepower potential over a given battlefield. The basic factors considered are accuracy, range, rate of fire, lethality, size and nature of targets, target priority, and tactical deployments. Essentially, force structure capabilities are calculated for a set of subdivisions of the battlefield. These capabilities are summed to obtain the resultant firepower measure.

b. This approach, which is adapted for use with digital computers, can be repeatedly applied for different tactical dispositions. Results of comparisons will yield valuable guidelines for research and development.

c. The model is useful as a means of determining optimum tactical deployments, as well as for determining weapons requirements.

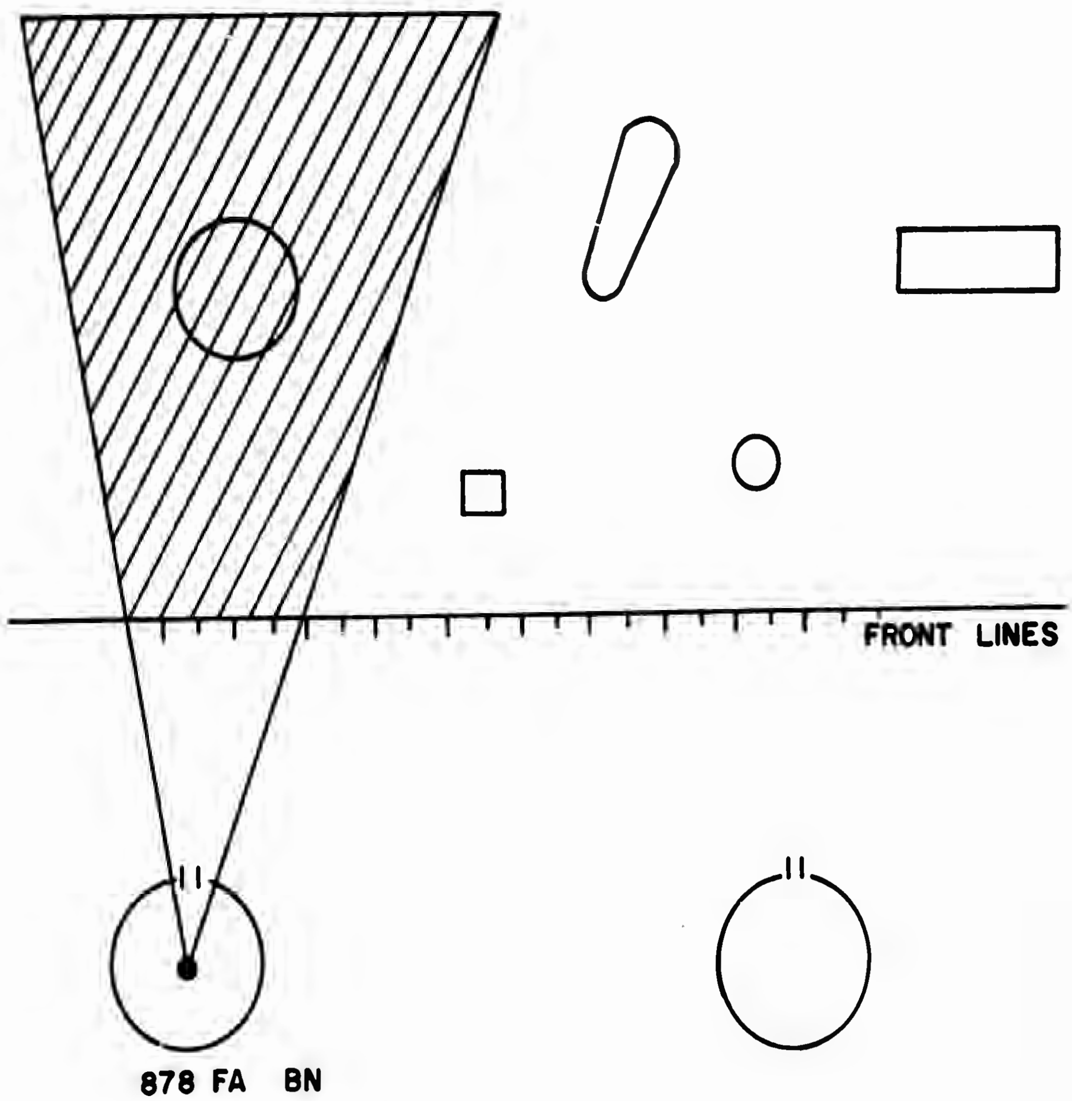
d. For actual application, it is necessary to extend and amplify the concepts and general outline presented here.

REFERENCES

1. "Effectiveness of Artillery Materiel in the Close Support Role" Project Nr USA A&GMS 56-12, 1 Feb 1957, Combat Development Department, USA Artillery and Guided Missile School, Fort Sill, Oklahoma.

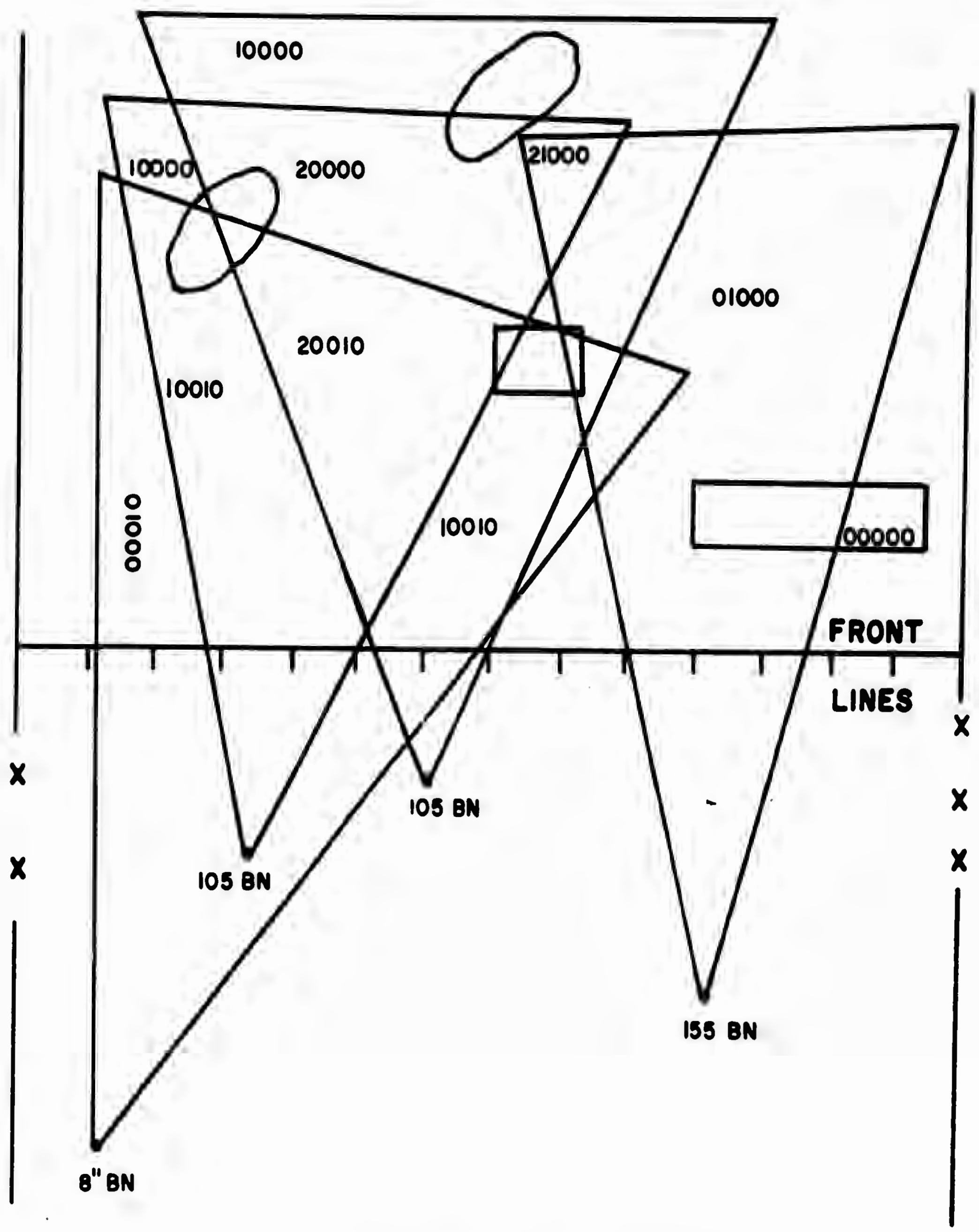
2. Staff Officer's Field Manual, "Nuclear Weapons Employment (U)," FM 101-31, Headquarters, Department of the Army, July 1959.

3. Cramer, Harald, _Mathematical Method of Statistics_, Princeton University Press, 1951.



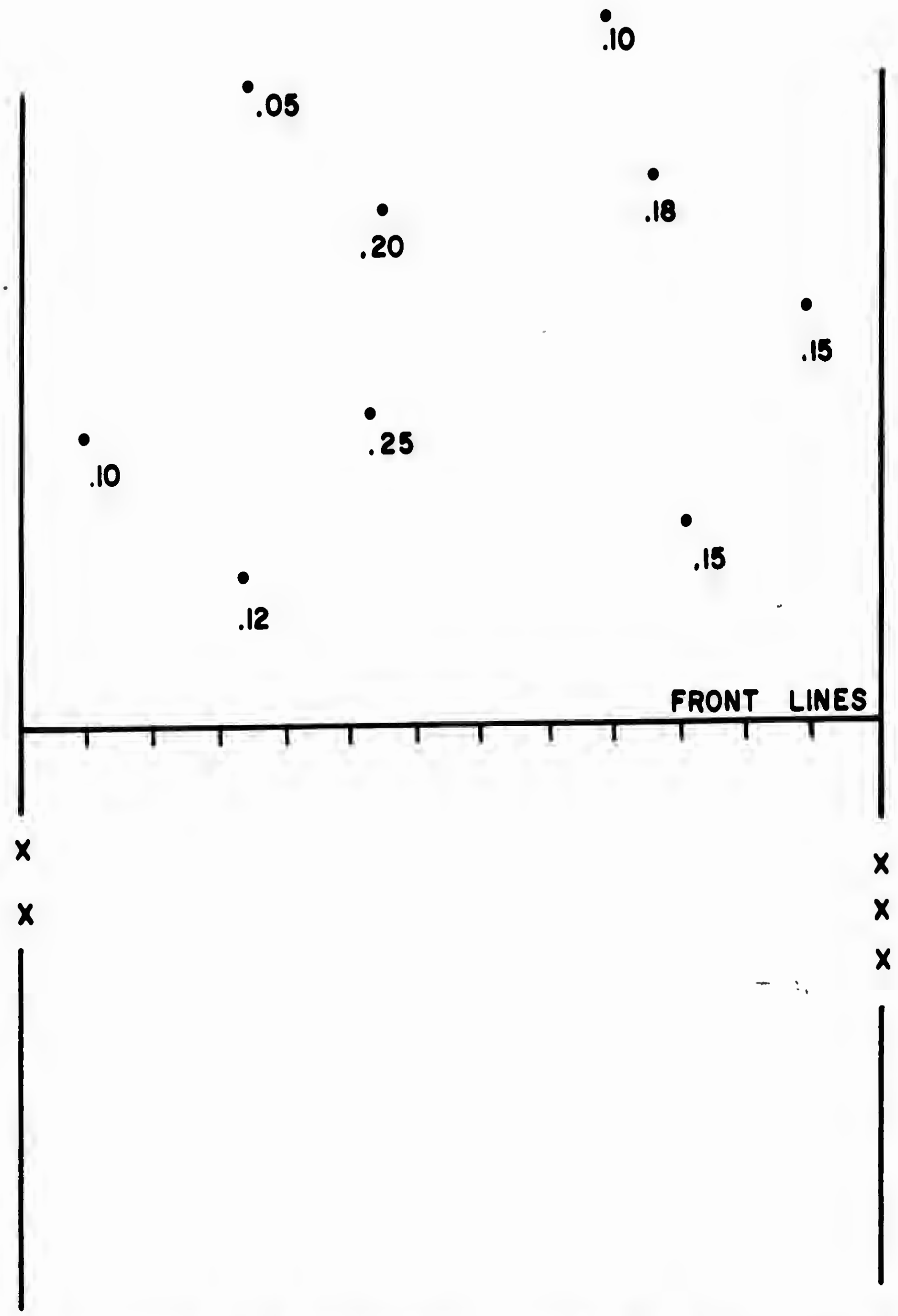
A CAPABILITIES CHART FOR A FIELD ARTILLERY BATTALION

FIG. 1



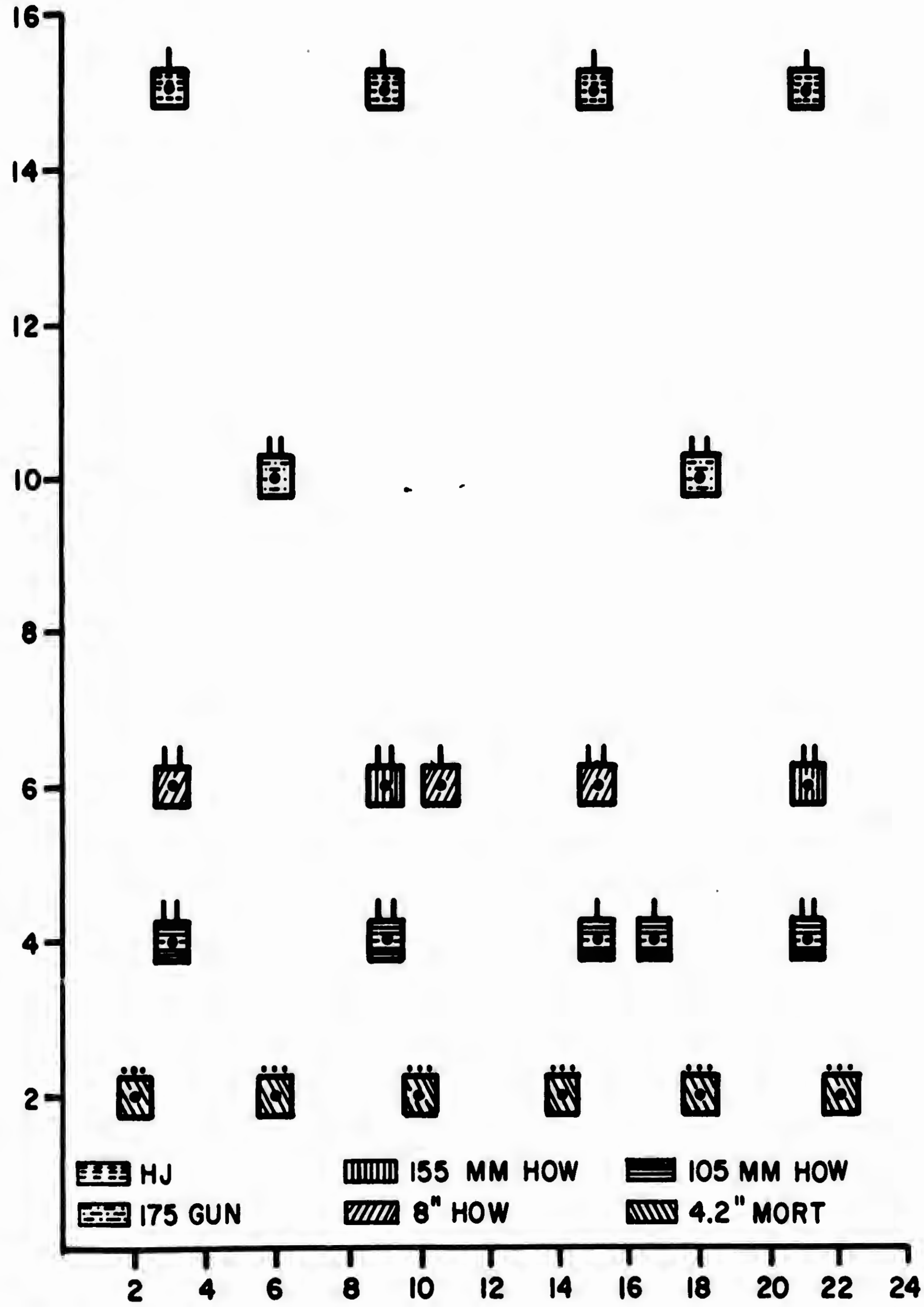
CAPABILITIES REGIONS
FOR SEVERAL WEAPON SYSTEMS
SHOWING DIFFERENT TYPES OF OVERLAP ZONES

FIG. 2



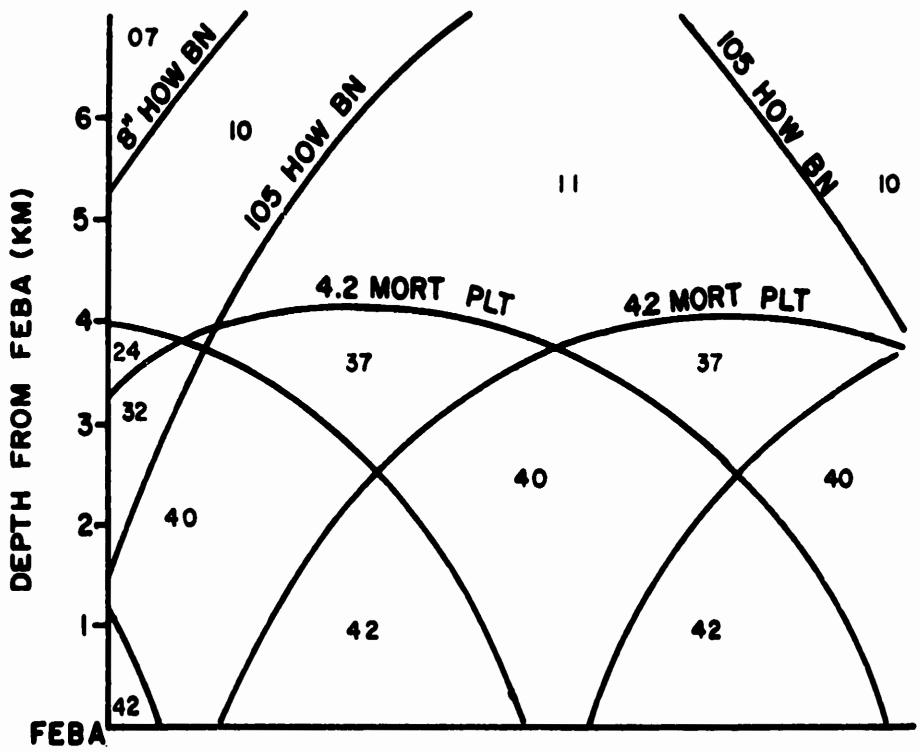
LOCATIONS OF DIFFERENT DAMAGE POTENTIALS OVER A BATTLE AREA

FIG. 3



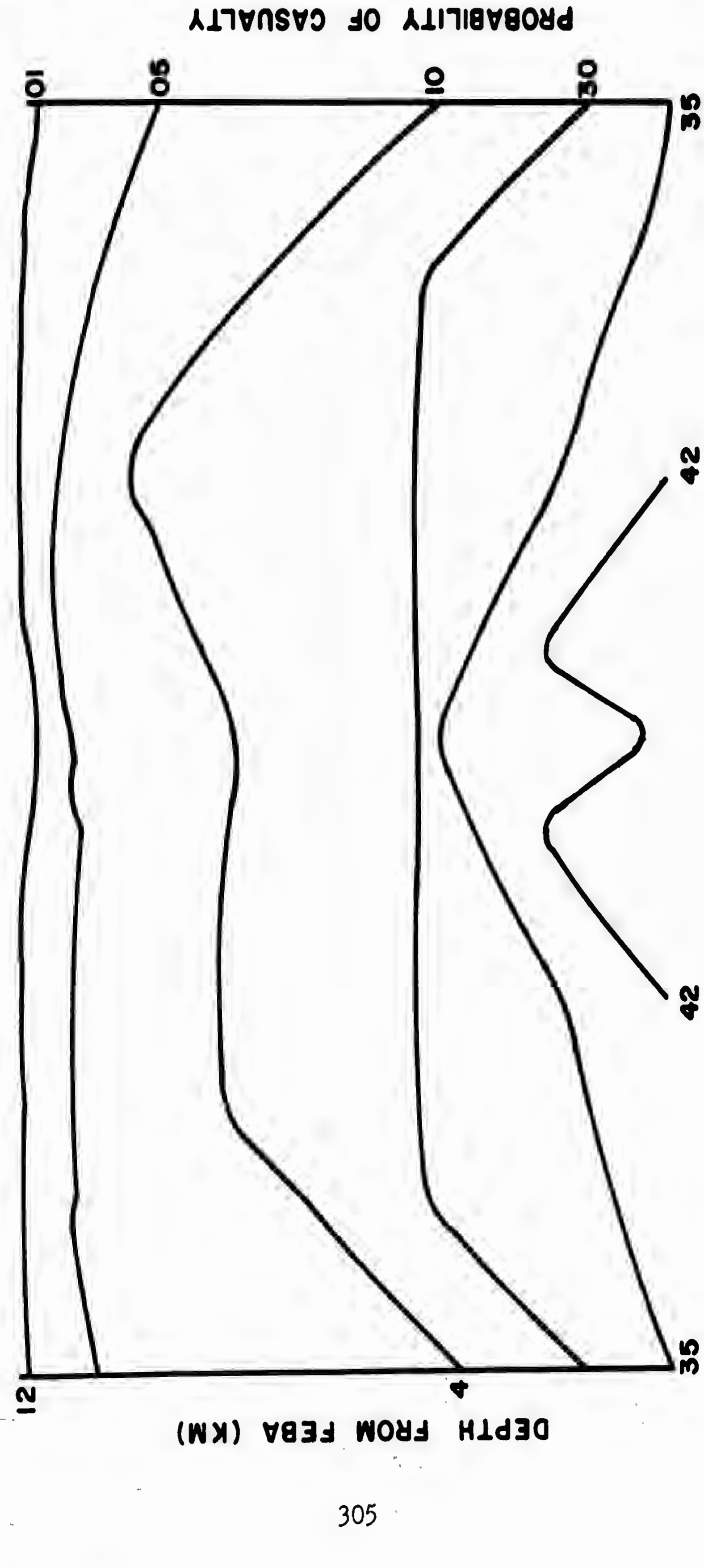
A TYPICAL ARTILLERY DEPLOYMENT

FIG. 4



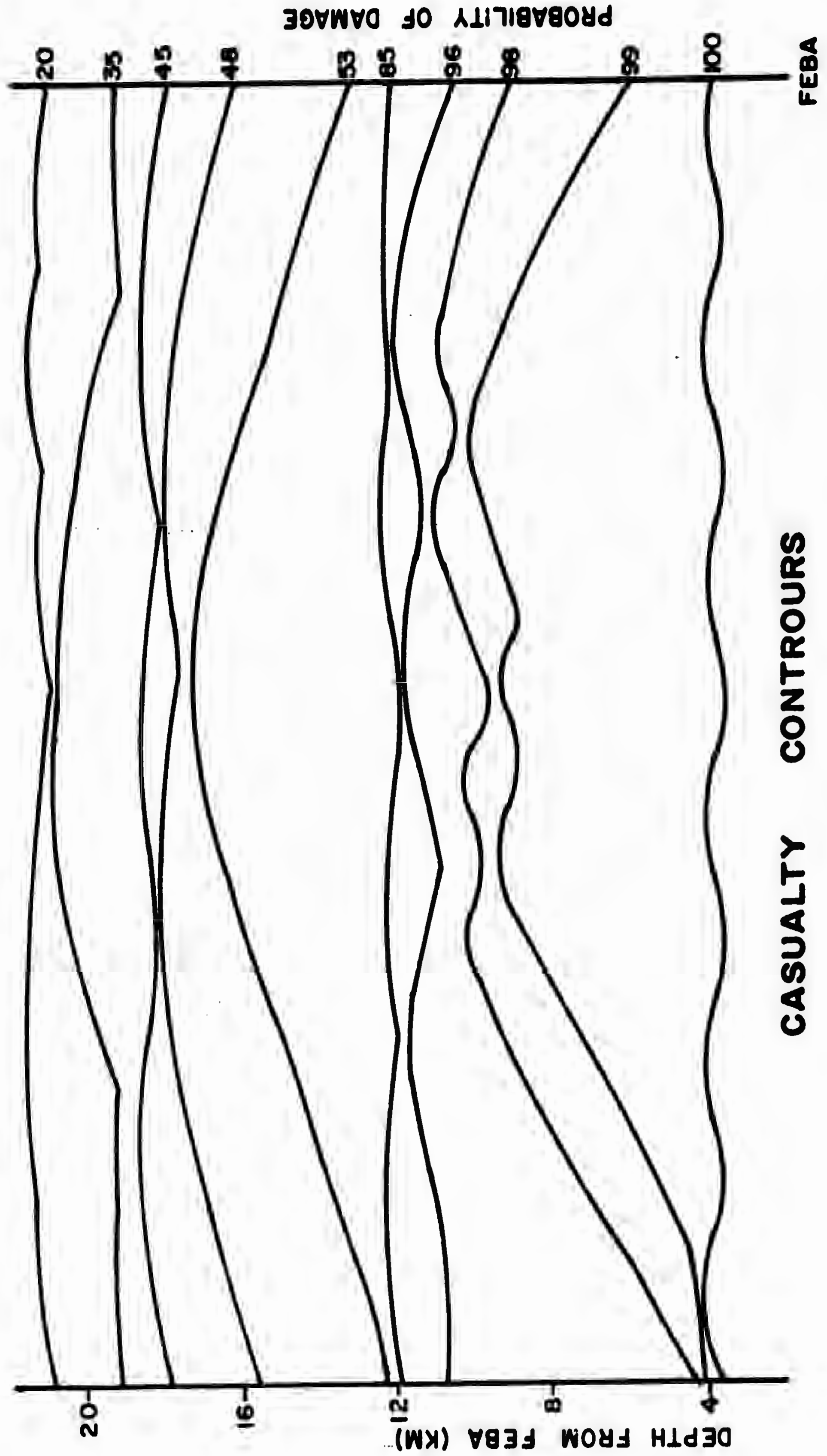
RANGE CAPABILITIES OVERLAP AREAS

FIG. 5



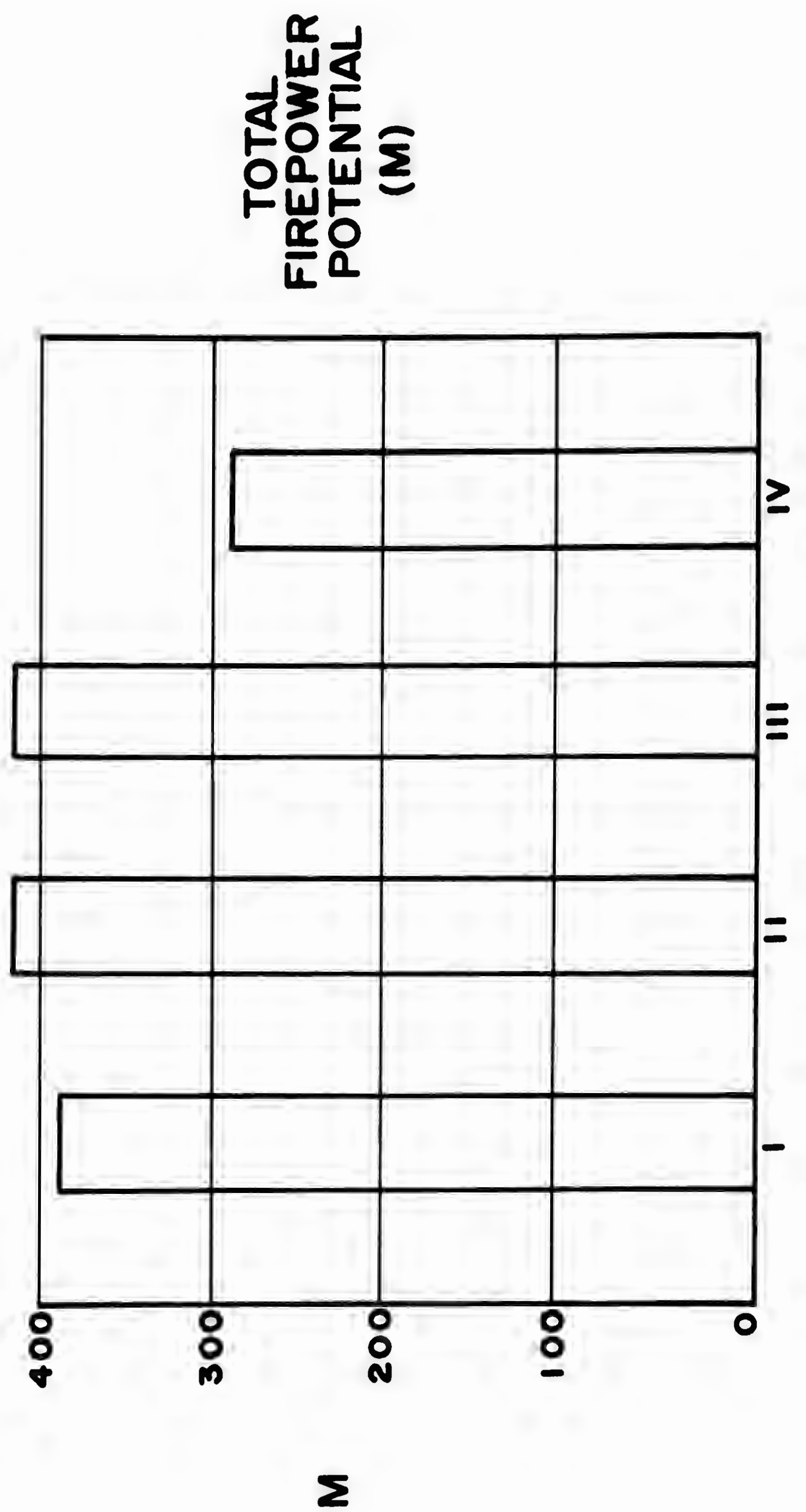
CASUALTY CONTOURS

FIG. 6



CASUALTY CONTROURS

FIG 7



DEPLOYMENT
FIG. 8

OVERALL RESULTS FOR MISSILES
25 KM SHORT OF FEBA

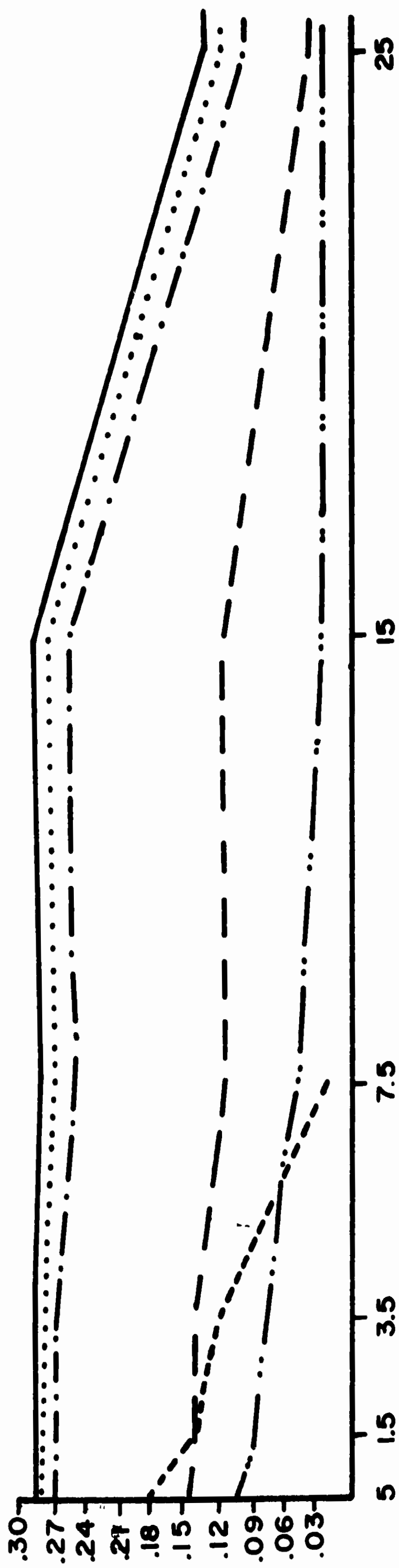


FIG. 9

EFFECTS OF AIR TURBULENCE UPON PROPAGATION OF LIGHT

HARALD W. STRAUB
DIAMOND ORDNANCE FUZE LABORATORIES
WASHINGTON 25, D. C.

INTRODUCTION

Several of the Technical Services have had difficulties in using photoelectric systems at near-ground level at distances of the order of miles between the transmitting and the receiving end. The disturbance generally observed was a random fluctuation of the received signal, although the transmitting light source was steady and unmodulated. On looking through the literature,[1] it appears that most observers, be they astronomers, optical or electronic scientists, have stumbled upon the same difficulties more or less independently of each other, each starting an investigation of his own. But the general consensus has been that the fluctuations are due to turbulence of the air through which the light propagated.

The investigations at DOFL were initiated for a somewhat different purpose; they were also independent, and the investigator at that time was also ignorant of earlier efforts. From an entirely different Ordnance project,[2] there had evolved the quasi-parallel-light concept[3] that permitted the construction of a rather unique spotlight - unique in that its beam pattern is completely non-divergent from the spotlight out to a point determined by controlled defocusing of the optical system. In the present case, the point was at a distance of 1300 feet from the spotlight; beyond 1300 feet, the spread was as small as 1 minute of arc.*

*The 1-minute-of-arc beam spread is comparable to that of a LASER beam.

It was known from theory that the $1/r^2$ law for the flux density does not hold at short distances from the spotlight, and the initial purpose of the work had simply been to ascertain experimentally at what distance from the spotlight the $1/r^2$ law would begin to hold. But it turned out that the measurements, which had to be made in the open countryside, were not as easy as originally anticipated. When a white cardboard screen was held in the beam at distances of 1.0 and 2.2 miles from the spotlight, the resulting light spots were far from uniformly bright. Quite to the contrary, what was observed were light spots which were traversed by smaller, black, areas of undefinable shape, 1 to 4 inches in size which, in turn, drifted in the direction of the prevailing breeze.** Because of the slowness of the drift, the intended photometric evaluation could not be performed with one of the short-time-constant meters.

INSTRUMENTATION

In order to obtain data which could be analyzed back in the laboratory, instrumentation was set up as shown in the self-explanatory Figure 1; this was done one hot night in September 1961 on one of the still unused runways of the Dulles International Airport under construction near Washington, D. C. Moving pictures of the shadow patterns were taken with air paths of 1.0 and 2.2 miles between the light source 1 and the condenser mirror 2 and at an elevation of 5 feet above ground. Since the built-in diaphragm stop in the camera 3 could not be used in this kind of a setup, the exposure was controlled by means of exchangeable neutral density filters 4 in front of the camera lens.

RESULTS

Figure 2 is a series of positive prints from frames which were taken at 1/16 of a second time intervals with exposures of 1/100 of a second at the 2.2-mile range.*** It shows how some typical spots appear to be drifting some distance across the face of the mirror from left to right, that is, in the direction of the prevailing breeze. This particular phenomenon has been called "frozen turbulence." The dark areas must be attributed to blobs of thinner-, or denser-than-average air, which have a life time of about one second and refract the light out of its original direction, possibly into adjacent bright areas.

** A similar phenomenon exhibiting horizontal striations instead of discrete spots was probably first described by Fawcett, Nature, 32, 224 (1888).

*** The movie from which these frames have been extracted will be shown in full at the Conference.

The dynamic range of the flux density was found to be about 1000 to 1 both at 1.0 and 2.2 miles. From this data, it can readily be understood that air turbulence can have a very disturbing effect not only upon a beam of unmodulated light, but especially upon a very short pulse of highly collimated light such as that from a LASER source.

There is another and equally disquieting phenomenon that has been observed while the moving picture was being taken: The entire beam was seen to exhibit a random motion, mostly in the vertical plane, but some lateral motion was also observed. At the 2.2 mile range, the vertical peak-to-peak amplitude amounted to about 18 feet (6 minutes of arc), with a periodicity varying between 3 and 15 minutes in time. As a matter of fact, it has not been possible in these experiments to determine the true, straight line of sight with any great certainty. The phenomenon is attributed to fluctuations in the overall density gradient of the near-ground air layers.

In order to observe the bending-beam phenomenon, it is obviously necessary that the spread of the beam be narrower than its deflection, and that the source be either CW-operated or at least pulsed more rapidly than the occurrence rate of variations caused by the atmosphere. It appears that such beams have not been available heretofore.

PLANS

Efforts are now underway at DOFL to perform the described experiments in a quantitative manner; to pin-point those portions along a turbulent light path which contribute the most to the perturbation; to determine the percentage loss of energy resulting from turbulence; to learn about non-linearities, if any, arising when higher and higher flux densities are used, such as from the high-energy LASERS which may be expected by 1965; and all of the above under as many different geographical, climatic, and meteorological conditions as are necessary to obtain a clear understanding of the perturbation phenomenon and its effects. However, even while these investigations progress, there have already been opportunities for direct application of the new knowledge on propagation to astronomy and to military systems such as range-finders and communication links, and more can be expected.

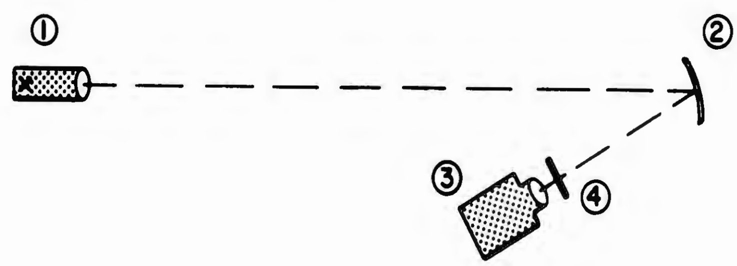
REFERENCES

(1) Bibliography: M. H. Wimbush, Optical Astronomical Seeing; A Review. Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii. Scientific Report No. 1. AF 19(604)-2292, May 1961. About 230 references.

(2) H. W. Straub, A Sensing System for Dynamic Armor (U). Report of Army Science Conference, United States Military Academy, West Point, New York. 26-28 June 1957. Vol. 4, p. 256-263. SECRET.

(3) H. W. Straub and J. M. Arthaber, The Quasi-Parallel Light Principle; Hollow-Conical or Flat Radiation or Detection Patterns (U). Proceedings of the 1959 Army Science Conference, United States Military Academy, West Point, New York. 24-26 June 1959. Vol. II, p. 407-414. UNCLASSIFIED.

INSTRUMENTATION FOR PHOTOGRAPHING TURBULENCE



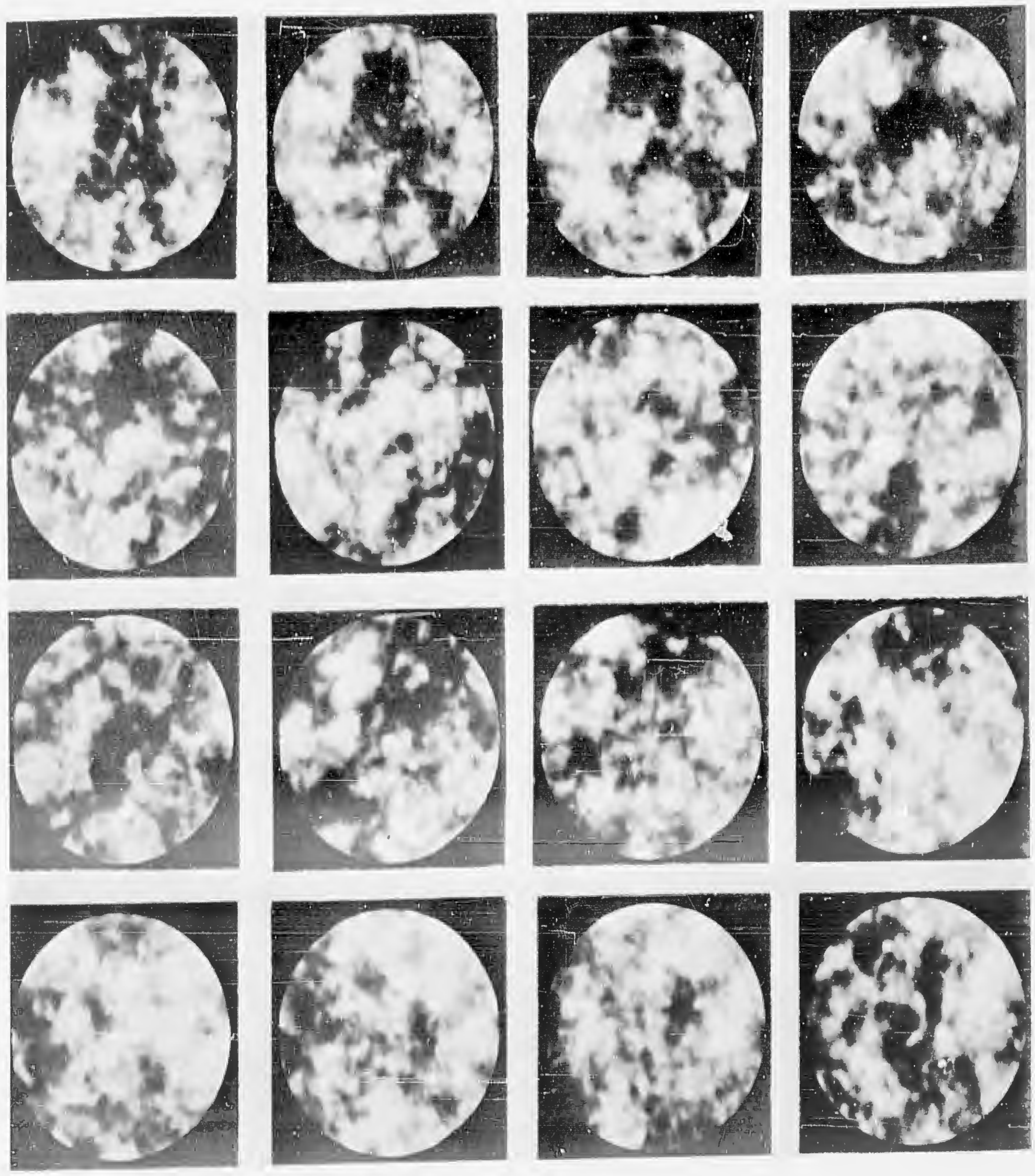
① QUASI-PARALLEL SPOTLIGHT•••as collimated "point" source

② CONDENSOR MIRROR•••images spotlight ① at objective lens of movie camera ③

③ MOVIE CAMERA•••focused on mirror ②

④ NEUTRAL GREY FILTER•••exchangeable to control exposure

FIGURE 1.



Mirror 2 of Fig. 1 with Superimposed Air Shadow Patterns. 16 Frames, 1/16 of a Second Apart, Taken over 2.2 Mile Air Path. Sequence Is From Left to Right and Down. Diameter of Mirror 12-$^{1}/2$ Inches, to Serve as Scale.

Figure 2

ACKNOWLEDGMENTS

The author wishes to thank Mr. John A. Dunbar, Mr. Donald W. Engle, Mrs. Flora Seckel and Mrs. Jean Green, of the Medical Audio-Visual Division at Walter Reed Army Institute of Research; Mr. Milton C. Rossoff, Medical Statistics Division, Office of The Surgeon General; and Mrs. Nancy Grigsby, US Army Medical Research and Development Command, for their help in the preparation of the Tables, Figures, statistics, and typing of this paper.

PROGRESS AND PROSPECTS IN IDIOPHYLAXIS
(Built-in Individual Self-Protection of the Combat Soldier)

MARION B. SULZBERGER, M. D., F.A.C.P.
US Army Medical Research and Development Command
Washington, D. C.

All who have been connected with any aspect of military research and development realize that weapons systems generally consist of three main parts: The man, the carrier or vehicle, and the weapon itself. Of these three, man is by far the most complex, most unfathomable and often the most fragile. And, although it may be heterodox to say so to an audience dealing with modern missile systems, tanks, transportation, communication and all kinds of very expensive hardware, I believe that man is also the most valuable military component, especially when trained and skilled in the various specialized crafts and intricacies of modern warfare. And, it is not just a matter of belief but one of record that in military campaigns from antiquity to Korea and Kuwait, the component man is the one that fails the most often. Moreover, he most often fails not because of bullets or missiles or any enemy action, but because of the stresses of climate and food and anxiety and disease.

Let me document this statement with the figures in this table based on the casualty lists of World War II.

(Table 1)

If these figures of World War II astonish you because the man-day losses due to diseases are four times as great as those due to battle injuries and wounds, the best available information indicates that the disproportion in this direction will be even greater in any future campaigns in the remote regions of the world.

We, in the medical service, who are responsible for the effectiveness of the component man are faced with Buck Rogers-Alice in Wonderland advances in technology; exponential expansions of scientific and medical knowledge; kaleidescopic rearrangements of policies and plans; rapid-fire introductions of new kinds of weapons, new kinds of transportation and communication, actual and potential dispersion

of combat and special forces throughout all the regions of the globe; and our troops' consequent unbelievably sudden and rapid encounters with new types of hostile climates and environments, new types of animals, poisons, vegetation, microorganisms, virus and other causes of disease. These new conditions bring with them constantly new and constantly waxing stresses upon the capacity of our men to perform with reliability and effectiveness. They also present constant challenges, as well as opportunities for medical and scientific research.

The soldier's reliability, his mental and physical health, cannot be maintained except by a specially-oriented military medical research and development program which is directly geared to every advance of science and at the same time rapidly responsive to the everchanging military needs. The statement, self-evident though it is in relation to other military research and development programs, such as those in Ordnance, Signal, Transportation, Engineering, and the like is not quite so self-evident and requires a bit of explanation when it is applied to medical research and development. For medicine--the science and the art of preventing, alleviating or curing disease and disability--would, on casual consideration, appear to be the same, whether it is to be applied to the man in or the man out of uniform. However, this is not the truth. All of us are kept fully aware of the billions of dollars which are being spent on civilian medical research; each of us is contributing to these out of his own pocket, not only in the form of taxes to support the programs of Federal, as well as State and municiple governmental agencies, but also in contributions to private hospital and medical school drives, marches of coins, cancer, heart, and mental retardation crusades, and so forth. But, despite the many vast and varied programs of civilian medical research, there remain medical problems which are not now and will not in the foreseeable future be investigated by any civilian agency with a vigor commensurate to the magnitude and immediacy of the threats they pose to our Army's successful operations. These are the medical problems which are the clear-cut obligation of the Army's Medical Research and Development Command.

I shall here recount just one example. In a recent test, 41 of our men were divided into two similar groups. One of these groups was trained, and in addition, acclimatized to heat while at Fort Knox in the United States. The other group was trained at Fort Knox at the same time and in precisely the same way but not heat-acclimatized. Both were then set down simultaneously in tropical Panama and put through the same military exercises and stresses. As had been expected and hoped, the heat-acclimatized men resisted the tropical heat and humidity much better than the men in the non-acclimatized group. In the 20 non-acclimatized men there were 10 casualties due to heat; while in the previously acclimatized group there were none. Moreover, the heat-acclimatized men were able to carry out a substantially greater amount of work in the heat. For example, in a load-

moving task during a 90-minute period the acclimatized group were able to move 11,200 pounds more of water than the non-acclimatized group. Table II shows the breakdown in pounds per man in each of the three 30-minute periods in which the two groups were moving the buckets of water.

(Table II)

But, in contrast to these expected and desirable results, an undesirable finding was that on the morning of the third day, 32 of the total of 41 men had blistering feet. The blistering occurred in 16 men in each group and was of equal severity in both groups; 4 men in each being incapacitated by the blisters.

Now blistering of the feet certainly affects both civilians and soldiers. But, while it is usually just a disagreeable and perhaps somewhat painful inconvenience to civilians, as a source of military incapacity severe blisters on the feet can result in the loss of a battle, a campaign, a cause. There is, therefore, an absolute mandate for military medical research to study intensively the factors and mechanisms which cause blistering--and how to prevent or minimize their effects in soldiers. This fact is supported by the figures of World War II showing that blistering dermatoses caused the loss of over 4 million man-days in the United States Army alone; and that in the British Malayan campaign in 1954-1955** the incidence of new cases of skin disease was about 1000 per thousand men per annum, and represented the largest single cause of admission to the hospital (46 per thousand per annum). At this point I would like to emphasize that in the present state of our lack of knowledge, research into the causes of blistering must go all the way from the most basic types of scientific, physico-chemical, biological, mycologic, bacteriologic, immunologic, and clinical research up to and including almost purely applied research and development in combination with other technical services, such as the efforts to discover and supply our men in the tropics with better boots, better socks, better formulas for foot powder, and generally better clothing and gear.

While it is true that incapacitating skin diseases, including blistering, are the fourth major cause of disability in our Armed Forces, I have cited them here not because of this but because they seem to me to be such convincing examples of conditions which are unequivocally military problems even though the same conditions affect also the civilian population.

In the diagram shown in Figure 1, I give my purely personal and approximate estimates of the probable relative military and

---

**Archer, T.C.R., Major, RAMC; Journal of the Royal Army Medical Corps, Vol. 104, No. 1, p. 1, January, 1958.

SCHEMATIC ESTIMATION OF RATIO OF MILITARY AND [illegible] REQUIREMENTS USEFULNESS AND APPLICATIONS OF BOTH BASIC AND APPLIED MEDICAL RESEARCH IN VARIOUS FIELDS AND DISEASES

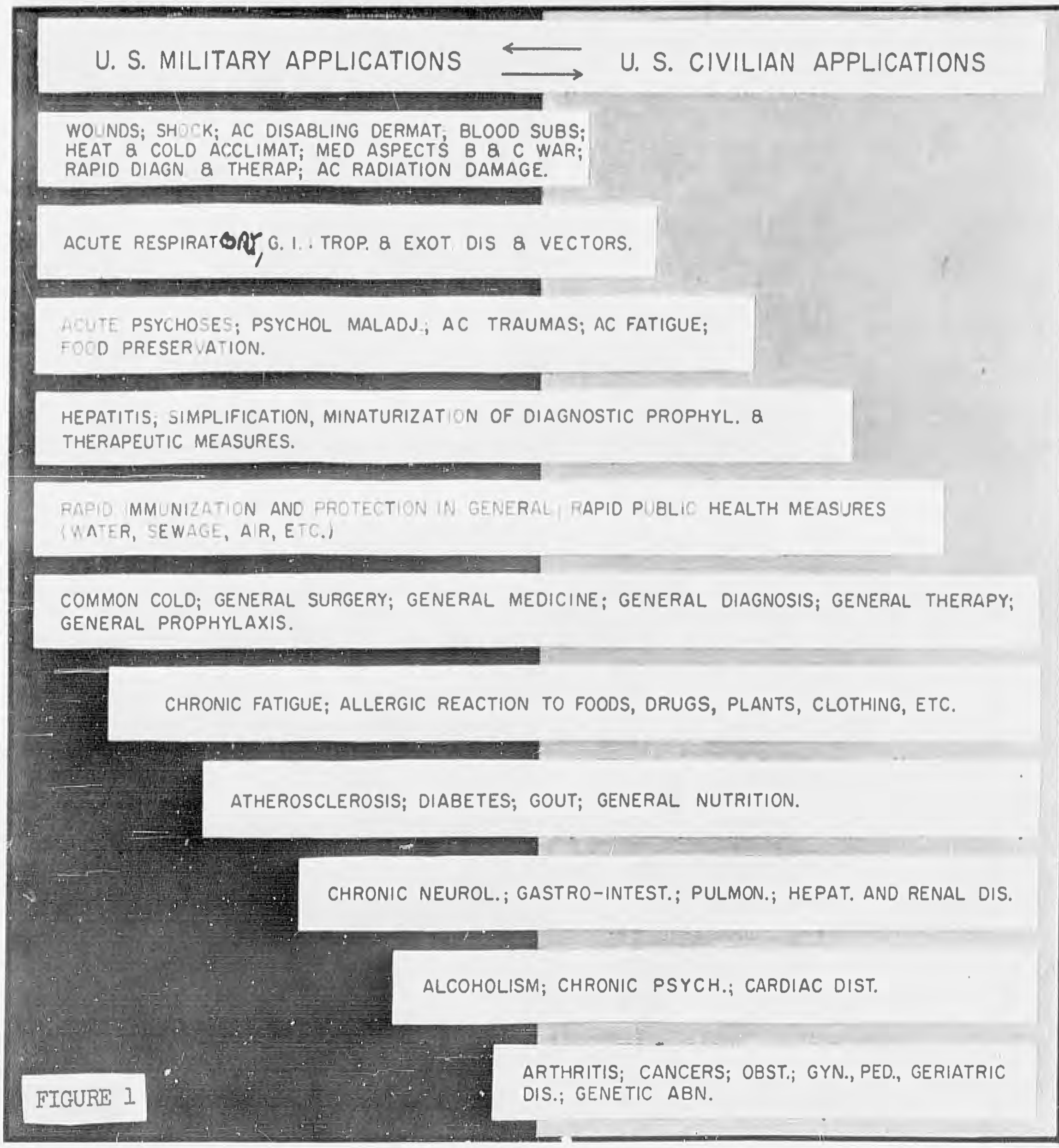


FIGURE 1

civilian importance of some other common medical problems. You will note that many problems that loom so large in the programs of non-military medical research do not figure prominently among those which I consider important to the United States Army (cancer, arthritis, chronic psychoses, alcoholism).

But the common problems which I have placed high on my list do constitute a most back-breaking and possibly disastrous burden in any future military operations. How great this burden in personnel and in logistics has been in the past is strikingly illustrated by the figures for which I am indebted to Colonel Vorder Bruegge, of our Command, as shown in Table III.

(Table III)

Perhaps the best way to bring about a substantial reduction in personnel and logistics requirements is, of course, by individual preventive medicine--by endowing each soldier entering a combat zone with his own in-built capability to protect himself against the prevalent diseases, stresses, climate and other on-slaughts.

The improvement and strengthening of the combat soldier's in-built self-protection by medical means is one of the central objectives upon which the United States Army's research and development program is focused. In discussing this program, Lt Col Donald Howie, Lt Col Jacques Sherman and I have given the name "idiophylaxis" to this form of protection--deriving the term from "idio" meaning self and "phylaxis" meaning protection, a neologism which is analagous to prophylaxis, anaphylaxis and similar terms in common usage in medicine.

Under idiophylaxis we do not include the protective devices which are supplied to the soldier to wear or to use in the form of such things as body armor, bullet-proof vests, anti-mine boots, water-repellent clothing, arctic suits, crash helmets, anti-flash goggles, insect netting, insect repellent creams, and so forth. We do include under idiophylaxis every form of protection that can be given to the soldier by preceding mental and physical preparation through medical means. Thus, idiophylaxis includes the mental conditioning, the immunizing procedures, the chemoprophylaxis, the medicaments and antibodies which we can cause to be embodied in the soldier's own person. It includes every protective capability with which we can medically endow him so that were he to be stripped suddenly naked, he still would carry substantial degrees of protection with him.

Mental conditioning, i.e., psychic idiophylaxis has been placed at the very top of my list because of the high priority which must be given to endeavors to equip our soldiers with the mind and the will to resist the terrible stresses which modern warfare brings with it. We must reduce his susceptibility to excessive fatigue and confusion, anxiety and mental breakdowns. Unfortunately, I shall not have

time to describe our very large and vigorous medical program aimed at a basic understanding of the complex functioning of man's mind in order to arrive at practical methods enabling him to support the impact of the war's sounds and sights, its darkness and its lonesomeness. But, I would like to emphasize that every protective measure with which we endow the soldier's body also contributes greatly to his mental idiophylaxis and his effectiveness and confidence. For, when the soldier feels the physical protection which we have been able to confer upon him coursing through his bloodstream or built into his own skin, he knows in his bones that everything has been done to protect him beforehand and everything will be done to help him if he gets into trouble later.

Another, and perhaps the most militarily important field of all, is the idiophylaxis which consists in conferring immunity or heightening resistance to various types of infectious diseases. This is going on continuously with unremitting vigor, and our global effort in this can be divided into three main phases:

1. Gathering from every part of the world information and specimens of diseases which may be encountered there by our soldiers.

2. Research on the causes and carriers of these diseases, their microorganisms and viruses and vectors in laboratory and experimental animals, cell cultures, etc., both here and abroad. And,

3. The endeavor to produce immunizing vaccines of all types and their laboratory tests and finally their clinical assays.

In these efforts the Army's Medical Research and Development Command collaborates closely with our sister services, with the National Institutes of Health, and with numerous other agencies, both on this continent and throughout the world. Moreover, we are lending major support to the Armed Forces Epidemiological Board and its Commissions, the Armed Forces Institute of Pathology, the National Research Council and similar agencies. In addition, many universities and medical schools, not only in the United States but elsewhere, are receiving our support and are assisting us in this program.

The following tables show some of the conditions in which idiophylaxis can now be conferred (Table IV); some in which chemoprophylaxis is possible at present (Table V); and some in which efforts are being made to increase our capabilities in both of these directions (Tables VI, VII and VIII).

These tables indicate not only the great progress which medicine has already made, but how much still remains to be done. And there remain, of course, the practical problems of how to inject so many different vaccines to large numbers of men in short periods of time with the utmost possible safety. In Figure 2, the injection

Injection Gun FIG. 2

gun, developed at the Medical Equipment Development Laboratory at Fort Totten, New York, is able to give immunizing injections to over 1000 individuals per hour, with accurately measured dosages, complete sterility and entire safety and elimination of the danger of communicating hepatitis or other infectious diseases from one infected individual to another. Moreover, with this gun, the personnel required on an immunizing team have been reduced to one half the number of those required with the usual syringe and needle method.

Here in this rather prognostic drawing of the idiophylactic soldier, Figure 3, I have indicated the immunizations and vaccinations against infectious diseases as having been given in the soldier's arm and buttock.

The other ways in which idiophylaxis is being conferred are also indicated, including the heat and cold acclimatization which we believe will now be possible simultaneously in the same person.

There are two points which I realized only after I had seen this finished drawing. One was that although it is to some degree a projection into the future, it does not appear to be nearly as fantastic as a drawing or description of our astronauts of today would have seemed to me only 15 years ago. The second point that I realized is that the majority of the protective capabilities indicated on this drawing have something to do with the protective capacities of the human skin. Perhaps this is why I, an old hand at dermatology, have been particularly interested in these problems. Not only does the skin serve as an accurate index and record of the immune capacities of the individual in the form of its many specific reactions to skin tests, but the skin is also a natural body armor protecting the individual in great measure against many of the on-slaughts of his environment. Thus, its sweat is one of the most efficient heat acclimatizing and cooling devices; its horny layer and melanin pigment among the best protectives against sunlight; its vascular system and its shunts one of the first defenses against cold and freezing; its surface lipid emulsion one of the best plasticizers and water shedders; the secretions of its surface one of the most powerful disinfectants and antibacterial agents; the barrier zone of its epidermis an almost perfect shield against penetration by most poisons. It would seem that the future soldier's increased protection or idiophylaxis would consist in some measure of a medically conferred augmentation of the already substantial inherent capacities of man's skin to resist.

There is reason for hope that this augmentation can be accomplished. For we are already able to give by mouth certain medicaments which become so firmly embodied in the surface horny layer of the skin that it will resist the propagation of disease-producing fungi for long periods of time. And with this principle established, other infections may perhaps be controlled by a similar mechanism.

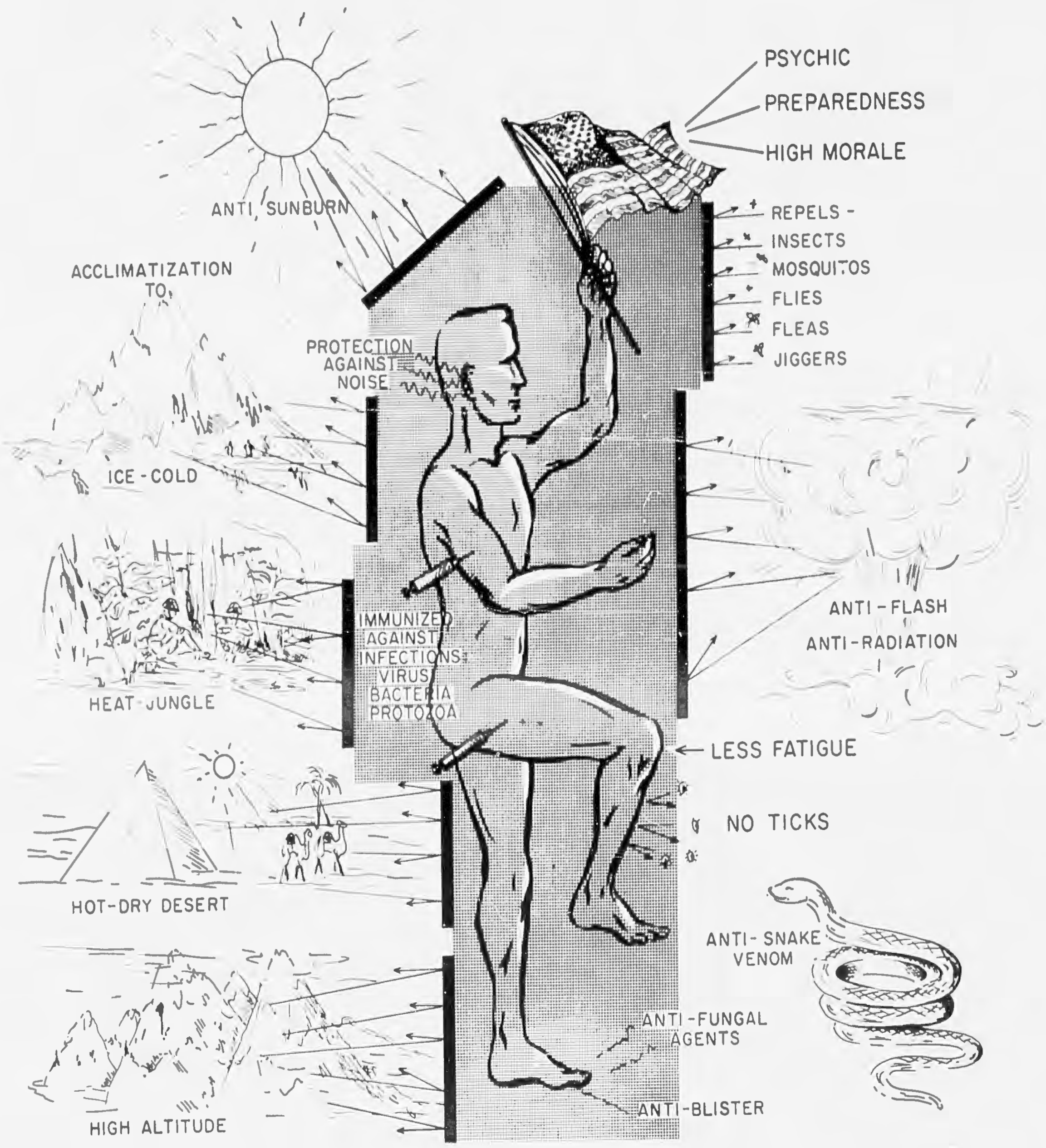


FIGURE 3

IDIOPHYLACTIC SOLDIER

Another promising field for idiophylaxis through the skin consists in the development of the soldier's self-protection against sunlight. While protective creams and protective lotions have long been available, these are difficult and messy to apply, sticky, wash off with sweat and rain, and are effective for very short periods of time. Our goal now is a chemical which will become incorporated in the keratin of the skin's surface and enhance its capacity to filter out the burning rays. There is already one such substance which can be given by mouth, but it has certain drawbacks. We believe that it may be possible to develop in the foreseeable future a better material which on internal administration or when sprayed on the skin's surface will become incorporated in the horny layer and produce a manifold increase in its sun and ultraviolet filtering capacity. Such a protective effect would then resist sweat, washing, rubbing off by clothing, and so forth for many days.

While the prospect is more remote and the problem much more difficult, there is some hope that we may even be able to develop chemicals with anti-thermal effects which will enable the surface of the skin to resist somewhat better than previously the rapid pulse of a flash thermal exposure.

Perhaps more important than any of these will be the development of a material which when incorporated in the skin either after local application or when taken by mouth, would make the skin's surface and its secretions repellent to insects, flies, mosquitoes, ticks, fleas, etc., and other vectors which are the bearers of the most important diseases affecting military operations, including malaria, yellow fever, sandfly fever, hemorrhagic fevers, and so forth. We are devoting a large effort to the discovery of such substances. We already know, for instance, that certain individuals carry with them in their own skin or on their skins' surface and its ingredients chemicals which either attract or repel such insects as biting mosquitoes, flies, bugs, and so forth. We know the exact chemical nature of a few of the most strongly attracting substances, and we have hopes, therefore, of finding some which will act as repellents and which the soldier can then produce in himself and on the skin's surface when given the adequate stimulus. All of you have known people who were particularly susceptible and who attracted insects; and other people who repelled them. It is through intensive study of the differences of the skin and its products in these two classes of individuals that we hope to make progress in the production of idiophylactic insect repellency. This does not, of course, mean that we are relaxing our efforts to discover ways of eliminating insect carriers of disease, or abating in the indoctrination of troops in the best methods of hygiene and personal measures of repellency while we are endeavoring to build into them the idiophylactic measures.

I have mentioned the prospects and efforts in idiophylaxis

against flashburns. While this surely will have to do with increasing the resistance of the skin and the skin's surface, there is a large program of endeavoring to develop better chemoprophylactic measures against the systemic effects of ionizing radiation, so that soldiers will be protected in sufficient measure to go into areas of fallout radiation which are still forbidden to the enemy. While there is no useful product as yet from this program, and none immediately in sight, I believe that the Army research efforts in this direction are well abreast of the most competitive efforts in this field anywhere in the world. Here, again, I must add that our research to confer chemoprophylactic idiophylaxis against radiation, against flash, and so forth, does not by any means interfere with our continued endeavors to discover and supply other protective measures, such as clothing, shelters, shielding and other items of equipment to ward off these damaging effects before they reach man or touch man's skin.

I cannot emphasize too strongly during this discussion of idiophylaxis that we believe that the summation of the effects of in-built self-protection plus all of the adequate equipment and protective devices which can be supplied to ward off damaging agents before they reach the individual will together offer the greatest promise and the greatest reduction of personnel and logistic burdens. While no one would be so naive as to think that research and development in the field of biological sciences and medicine would require the same magnitude of funding as the development of a large intercontinental ballistic missile system, a new type of super bomber, a nuclear submarine, and so forth, we do believe that a substantial increase and a substantially expanded military medical research and development program is needed, and the best possible investment, and that these facts are inescapable conclusions from the data which I have just presented.

In closing, I would like to assure you that we in the medical service are not so starry-eyed as to think that we will ever be able to confer upon the soldier a degree of idiophylaxis which will protect him against all of the attacks of nature or of an enemy, or to make his skin so tough that it will ward off bullets and flames, blast and all radiation effects. However, we do believe that we must develop the idiophylaxis of our soldiers to the utmost degree possible, and that we have every prospect of making him in this way the most effective and most resistant of all human beings and thus reduce the vulnerability of the most delicate component of our weapons systems--the trained man. If we succeed in this, we will <u>all</u> be rewarded, not only by the production of a more effective Army with the highest possible morale, but also by the happy medical yield of discoveries which will contribute to the reduction of disease and suffering, not only in soldiers, but in all mankind.

What this may mean for the health, welfare and strength of our nation has been well expressed in the opening sentence of

SULZBERGER

President Kennedy's message to Congress on February 27th, 1962, just a day after I had submitted my outline for this presentation. "The basic resource of a nation is its people. Its strength can be no greater than the health and vitality of its population. Preventable sickness, disability, and physical or mental incapacity are matters of both individual and national concern."

TABLE I

Personnel Losses from Disease and Battle Injuries or Wounds
(Based on Average Strength of 6,300,000 for US Army in WW II)

Average Number of Noneffectives each Day

Disease	190,386
Battle Injuries or Wounds	48,699
Total	239,085

Equals a Loss of

14 full-strength divisions

(11 because of disease; 3 because of battle injuries or wounds)

TABLE II

LOAD-MOVING TASK PERFORMED BY HEAT-ACCLIMATIZED GROUP AND NON-HEAT-ACCLIMATIZED GROUP IN POUNDS PER MAN			
Group	Periods		
	1st 30 Min.	2nd 30 Min.	3rd 30 Min.
Acclimatized	920 lbs.	1110 lbs.	1130 lbs.
Non-acclimatized	770 lbs.	900 lbs.	720 lbs.

(N.B. Both groups improved in second period through learning, but only acclimatized group sustained improvement through third period.)

TABLE III

Costs and Logistical Burden of Disease, Nonbattle Injury, and Battle Injury or Wound
(Calculations Based on Figures of WW II)

500,000 Troops Overseas (Require within 90 days)

25,000 Hospital Beds

28 Engineer Battalions or

21,000 Engineer Troops

working 1.5 months

86,000 Short Tons Material

28,500 AMEDS Personnel

2,100 MD's

23,100 KW Power

3,000,000 Gallons Water Per Day

TABLE IV

Idiophylaxis Now Conferrable by Immunologic Measures (Vaccines, etc.)

Adenovirus Resp. Inf.	Paratyphoid A & B	Rocky Mt Spotted Fever
Cholera	Plague	Smallpox
Diphtheria	Measles	Tetanus
Epidemic Typhus	Mumps	Tuberculosis (Probably)
Infantile	Q Fever	Typhoid
Influenza	Rabies	Whooping Cough
	Yellow Fever	

TABLE V

Idiophylaxis Now Conferrable by Chemoprophylaxis (Drugs)

Gonorrhea		Q-Fever
Malaria		Rheumatic Fever
	Syphilis	

TABLE VI

Idiophylaxis May Soon Be Available

Anthrax	Rift Valley Fever
Botulism	Russian Tick-Borne Fever
Encephalitides (By certain virus)	Tularemia
Leptospirosis	Heat, Humidity, Cold and Altitude Effects

TABLE VII

Some Examples in Which Idiophylaxis is Required (Perhaps Possible but not yet Available)

"Common Cold"	Epidemic Hemorrhagic Fevers (various)
Chikungunya Fever	Fungus Infection (various) (superficial and deep)
Dengue	Leprosy
Diarrheal and Enteric Diseases	O'Nyong-Nyong Fever
Encephalitides (various)	Relapsing Fevers (Tick-Borne and Louse-Borne)

TABLE VIII

Some Examples in Which Idiophylaxis is Required (Perhaps Possible but not yet Available)

(Continued)

Staph and Strep Infections	Sleeping Sickness
Wound Infections Chlostridium Pseudomonas and others	Radiation Damage
	Thermal Flash Damage
Parasitic Diseases (Miscellaneous, worms, flukes, etc.)	Fatigue
	Battle Psychoses
Schistosomiasis	

GROWTH OF HEXAGONAL FERRITE SINGLE CRYSTALS FOR APPLICATION AT MICROWAVE & MILLIMETER-WAVE FREQUENCIES

ARTHUR TAUBER and R. O. SAVAGE, JR.
U. S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY
FORT MONMOUTH, NEW JERSEY

INTRODUCTION: Today as never before great needs exist for secure, reliable, multichanneled, long distance detection and communication systems. The microwave region of the spectrum has been under intensive investigation for many years. The fruition of this research and development are the familiar radar detection systems, indispensable to our nation's security, and the long-range telephonic and video microwave communication networks. The millimeter-wave spectrum is particularly inviting for communications from satellites and astro-systems because of its narrow beam and large band. Furthermore, this lower frequency of the electromagnetic spectrum has become so crowded that one is literally forced to investigate higher frequencies for purposes of communication. Communication in the millimeter range is very attractive because thousands of messages may be transmitted by a single system with a very high level of security.

This paper describes the growth of new hexagonal ferrimagnetic crystals for standard and unique device applications at microwave and millimeterwave frequencies. The crystals are synthesized by recrystallization from molten fluxes in platinum crucibles. Three different flux systems will be described. Intrinsic thermomagnetization properties have been studied and some novel properties found. New theoretical models of magnetization are required to account for these properties. Microwave-millimeter-wave measurements of ferromagnetic line width demonstrate the low losses of some of these materials at frequencies between 9.5 and 70 kmc/sec. (kilomegacycles/sec).

The hexagonal ferrites, therefore, find standard applications as isolators, rotators, attenuators, etc., at frequencies where other materials operate inefficiently or for which no other materials are available. Non-linear devices such as power limiters and harmonic generators (frequency doublers) are particularly interesting, making possible high-power devices and extending our range of operation into the millimeter-wave spectrum.

Hexagonal ferrites have the additional attractive property of being made into permanent magnets, so that no d-c biasing field would be required on a device. In the present state of the microwave art permanent magnets weighing several pounds are required. Doing away with these magnets would result in ecomony as well as reduction in size and weight.

UNIAXIAL AND PLANAR CRYSTALS: The hexagonal ferrites to be described in this paper differ from the well known cubic (spinel) ferrites in that they possess a distinctive easy direction of magnetization (large anisotropy). Anisotropic magnetization is the origin of the unique properties and the applications of these crystals.

Hexagonal ferrites may be divided into two groups: uniaxial with an easy axis of magnetization and planar with an easy plane of magnetization. The distinguishing properties are demonstrated in Figure 1. The crystal adhering to the left leg of the magnet orients itself so that its c axis lies along the strongest magnetic field vector. This is the uniaxial type. The crystal adhering to the right leg of the magnet orients itself so that the plane perpendicular to the c axis (the plane containing the a and b axes) lies along the strongest magnetic field vector. This is the planar type.

CHEMICAL CONSTITUTION AND STRUCTURE: All of the hexagonal ferrites are structurally and chemically related to barium ferrite ($BaFe_{12}O_{19}$). They differ from barium ferrite in that they contain two divalent cations per molecule of the following metals: Mn, Fe, Co, Ni, Cu, Zn and Mg. Table I lists all the known types (1).

TABLE I

Hexagonal Ferrites

Chemical Composition	Symbol	No. of Oxygen Layers	Lattice Parameter c axis in Å
$BaFe_{12}O_{19}$	M*	2x5	10x2.32=23.2
$Ba_2Me_2Fe_{12}O_{22}$	Y+	3x6	18x2.32=41.8
$BaMe_2Fe_{16}O_{27}$	W	2x7	14x2.32=32.5
$Ba_3Me_2Fe_{24}O_{41}$	Z	2x11	22x2.32=51.0
$Ba_2Me_2Fe_{28}O_{46}$	M_2S	3x12	36x2.32=63.5
$Ba_4Me_2Fe_{36}O_{60}$	M_2Y	3x16	48x2.32=111.4

*Nomenclature devised by Smit and Wijn (11).
+All of these compounds are planar

All combinations of types and divalent cation are possible compounds. The symbols in the second column from the left will be used as a short-hand notation in refering to these compounds.

About twenty years ago, Adelskold (2) showed that barium ferrite had a hexagonal crystal structure. It consists of close-packed oxygen layers with metal cations in spaces between oxygen atoms. There are two kinds of layers and these are repeated alternately five times until the smallest repeating pattern emerges. If, as indicated in Table I, each layer is approximately 2.32Å (angstrom units) high, a unit cell of ten layers is 23.2Å high. The a axis of this material is 5.9Å long. The crystal symmetry of this as well as the other hexagonal ferrites accounts in large measure for the presence of the large magnetic anisotropy. Barium ferrite is quite useful as a ceramic permanent magnet in many applications (3).

The newer hexagonal ferrites are derived from barium ferrite by adding more oxide layers and forming slightly different blocks of atoms. This may be confirmed from an examination of the third column of Table I. One notes either 2 layers (compounds with true hexagonal symmetry) or 3 layers (compounds with rhombohedral symmetry). The length of the a axis is essentially the same for all compounds and equal to that of barium ferrite. All compounds with the Y-type structure have easy planes of magnetization and are designated planar. All cobalt compounds, regardless of type, are also planar.

SINGLE-CRYSTAL GROWTH OF BARIUM FERRITE: The first reported effort to prepare single crystals was by Wijn (4) who tried to grow barium ferrite from a mixture of barium oxide and magnetite. He did not succeed, but he did make a very important discovery which we shall discuss shortly. In 1958, Kooy (5), Brixner (6), and Mones and Banks (7) all reported methods of growing $BaFe_{12}O_{19}$. They employed growth from a melt contained in a platinum crucible using a flux. Of the three, the one described by Mones and Banks appeared most desirable because it best met the following three criteria of flux growth:

1. The flux depresses the temperature at which growth can take place.
2. The flux does not attack the crucible and is easily removed with a suitable solvent.
3. The flux yields crystals which are pure, of adequate size, and sufficiently perfect.

The method of Kooy's fails on 1 and 2, while the method of Brixner fails on 3. Using Na_2CO_3, crystals could be grown from a temperature of 1250°C. This flux does not attack the crucible and is easily removed with water or dilute acid. Gambino and Leonhard (8), of our Laboratory, prepared large crystals, two inches on edge, and chemically pure.

Trivalent iron (Fe^{3+}) may be replaced by trivalent aluminum in barium ferrite on a one-for-one atomic basis, up to 75% of the total iron content (7). DuPre, et al, (9) have shown that the magnetic anisotropy of this compound increases with increasing

aluminum substitution. An increase in anisotropy, as will be shown, extends the useful range of this material to millimeter-wave frequencies. Part of the present work was the growth of single crystals of barium ferrite containing aluminum substitution. The method employed was that of Gambino and Leonhard (8). Large, excellent quality crystals with a maximum of 12% iron replaced by aluminum were obtained by this Na_2CO_3 flux method. Attempts to prepare crystals with higher concentrations of aluminum by the previously described techniques were unsuccessful.

GROWTH OF NEW HEXAGONAL FERRITES: Wijn's attempt to grow barium ferrite was described. Instead of this compound he obtained several of the compounds listed in Table I. His further efforts to prepare these as single crystals by direct melts led to variable compositions and undesirable intergrowth of different crystals. By 1959, a large list of compounds of this new type and some of their intrinsic properties were described by Smit and Wijn in a book entitled "Ferrites" (11).

The work of Wijn demonstrated that these compounds decomposed upon melting in air, just as barium ferrite. This is frequently the case with iron-containing compounds. Consequently, the flux method was again considered for crystal growth. Because of the similarity between the chemistry and structure of these compounds and barium ferrite, Na_2CO_3 was considered as a flux. Simple chemical considerations show that in such a complex system, Na_2CO_3 could not be the equilibrium compound acting as a flux. Knick and Kohlmeyer (10) indicate that the equilibrium compound in the system Na_2CO_3-Fe_2O_3 in air was $NaFeO_2$. Realizing that this was probably the flux in the reactions to be considered, we attempted to prepare this compound. In every instance, through a wide temperature range, Na_2FeO_4 was the equilibrium compound obtained. These experiments indicate that earlier work in the literature on this system requires considerable revision.

Over 200 crystallization experiments have been carried out in the general system, Na_2FeO_4-Fe_2O_3-BaO-MeO, where MeO stands for a divalent oxide. Na_2FeO_4 was recovered as the equilibrium sodium-iron-oxide compound in these experiments, confirming our earlier work. All of the crystal types listed in Table I have been grown with zinc and cobalt as the divalent cation. In addition, some of these types have been grown with nickel, magnesium, and copper as divalent cations. Compounds with manganese cannot be grown from this flux system, because this element tends to be oxidized to the hexavalent state in the alkaline environment associated with the presence of sodium. Manganese compounds have been grown from a system containing $Mn_2Fe_2B_2O_{10}$ as a flux. Lead oxide has recently been used quite successfully as a flux to grow many interesting magnetic oxide crystals. This cannot be used as a flux for the compounds considered here because the lead would replace barium. Bismuth trioxide (Bi_2O_3)

is sometimes used to grow crystals and it has been used in this very successfully. It is, however, presently limited to W type crystals because a competing side reaction leads to the formation of cubic $BaFeO_3$ crystals.

EXPERIMENTAL PROCEDURE IN CRYSTAL GROWTH: The general procedure is to melt a stoichiometric mixture of reactants (oxides or carbonates of the appropriate metal) in a platinum crucible, heating in an electric furnace, at a temperature between 1100° and 1375°C. After sufficient time to insure solution, the furnace is automatically cooled at 0.75° to 4°C per hour with close control of temperature (±2°C), to a temperature at which the contents of the crucible are solid. The furnace is then cooled more rapidly to approximately 600°C, where the crucible is removed from the furnace and cooled to room temperature. Flux is dissolved away and the crystals are freed. Spectrochemical analysis is performed to determine chemical purity. X-ray diffraction patterns of the crystals are made to determine the type of crystals, presence of intergrowth or twinning, and lattice parameters.

All systems from which crystals have been grown consist of four components. The principal system investigated was Na_2FeO_4-Fe_2O_3-BaO-MeO. Little difference has been found between the systems in which Me is cobalt or zinc. Figure 2 is a perspective quaternary diagram in which zinc is the divalent cation. The quaternary diagram is a tetrahedron along whose axes the compositions of the four components are represented. Thus, stability regions of various compounds may be given in terms of the concentration of the four components. All of the compounds that are known to exist in the quaternary system are indicated. It is quite difficult to represent compositions in such a diagram without misleading distortion. Fortunately, two components, BaO and ZnO may be held fixed and the remaining two, Na_2FeO_4 and Fe_2O_3, varied. We may now represent compositions in terms of families of planes penetrating the tetrahedron. One such plane, the BaO:ZnO 1:1 mole fraction, is shown hatched in Figure 2. The experimental results for the BaO:ZnO 1:1 plane are shown in Figure 3. The findings for the corresponding plane where Co is the divalent cation are essentially the same. In Figure 3 the composition of Fe_2O_3 is indicated along the abscissa and that of Na_2FeO_4 along the ordinate. The circles represent experimental runs which were not necessarily at equilibrium. Interestingly four of the five compounds crystallize in the 1:1 plane. This probably arises because the stoichiometric compositions cluster close to one another near the center of the composition diagram where the 1:1 plane lies. The dashed lines, which are somewhat arbitrary, mark the single-crystal composition fields. It should also be noted that the single-crystal fields are quite narrow. This too is a consequence of the clustering of the stoichiometric compositions.

In addition to the 1:1 composition planes, those corresponding to BaO:ZnO or BaO:CoO ratios of 2:1, 3:2 and 1:2 were investigated. The single crystal with a particular BaO:MeO ratio always occurred in its plane. However, one or more of the other crystal types always appear in that plane with the exception of the 1:2 plane. These results are given in Table II and apply also to Co or Zn compounds. Several other planes were explored, and data from these are included. The phases with the largest stability regions were W and Y.

TABLE II
Crystallization in the Quaternary System
Na_2FeO_4-Fe_2O_3-BaO-MeO

BaO:MeO Plane From Which Grown	Hexagonal Crystal Type Observed: Principal	Additional
1:1	Y, M_2S	W, M_2Y
1:2	W	
3:2	Z	Y, M_2S
2:1	M_2Y	M, Y
1:6	---	M, W
1:8	---	M, W
7:6	---	M, Y, W, M_2S
5:4	---	M, Y, W, M_2S

CRYSTALLINE MORPHOLOGICAL PROPERTIES: Single crystals up to $\frac{1}{2}$-inch on edge have been obtained by the flux method. All the crystals grown in this study are characteristically black with highly reflecting faces, Figure 4. (Zn)Y crystals are generally tabular, Figure 4. Density measurements on 20 (Zn)Y crystals yielded an average value of $5.46 \pm .04$ g/cm^3 (theoretical density 5.46 gm/cm^3). This together with spectrochemical analysis (purity = 99.9%) attest to the purity and quality of the crystals.

ELECTRICAL RESISTIVITY: The electrical resistivity of these crystals was measured with two probes on a high-impedance resistance bridge. These measurements yielded average resistivities of 100 meg ohms-cm. The crudeness of the measurement allowed only this order-of-magnitude result. Nevertheless, it was adequate to indicate that high-frequency applications were possible with negligible eddy-current losses. It also demonstrated that the crystals were free of even small amounts of ferrous (Fe^{2+}) iron, the presence of which decreases the resistivity markedly.

INTRINSIC THERMOMAGNETIC PROPERTIES: Magnetization as a function of temperature has been studied for many of the hexagonal ferrites in polycrystalline form (11). Differences between polycrystalline bodies and single crystals are due to such variables as orientation, density, homogeneity, and impurities which can never be as well controlled in polycrystalline materials as in single crystals. The difference becomes dramatically revealed when ferromagnetic resonance variables and microwave devices are examined.

The saturation magnetization of a uniaxial (Zn)W crystal is dotted as a function of temperature in Figure 5. The Y type crystal has an easy plane of magnetization. There are considerable differences between the (Zn)Y single crystal magnetization reported in this study and that given by Smit and Wijn (11) on a polycrystalline sample. The magnetic moment per formula unit, obtained by extrapolating the saturation magnetization curve to absolute zero and infinite field, is 18.9 Bohr magnetons compared with 18.4 obtained with polycrystalline samples (theoretical = 20 Bohr magnetons). The anisotropy field, measured on single crystals, was 9900 oersteds compared with 9000 obtained with oriented polycrystalline bodies. The ferromagnetic resonance line width, to be considered in another section, is 30 times as narrow in single crystals as in polycrystalline specimens.

The data in Figure 5 have not been previously reported even for polycrystalline samples. The unsaturated magnetization curve for the crystal oriented in the hard direction is also plotted. It should be noted that a reversal in anisotropy takes place at about 475°K. This suggests that (Zn)W undergoes a transformation from an easy axis to an easy plane above this temperature. Such a transformation has not been previously observed in these hexagonal compounds. Magnetic data for these as well as other single crystals are given in Table III.

TABLE III
Intrinsic Magnetic Data for Hexagonal Ferrites

Compound	Symbol	Saturation Magnetization at 20°C, $4\pi M_s$ in gauss	Magnetic Moment in Bohr Magnetons Experim	Theor
$BaFe_{12}O_{19}$	M	4800	20.1	20
$BaFe_{11.5}Al_{.5}O_{19}$	M	3900	17.0	17.5
$Ba_2Zn_2Fe_{12}O_{22}$	(Zn)Y	2600	18.9	20
$BaZn_2Fe_{16}O_{27}$	(Zn)W	5530	38.2	40.0
$BaCo_2Fe_{16}O_{27}$	(Co)W	4870	----	44
$Ba_4Co_2Fe_{36}O_{60}$	(Co)M_2Y	5020	----	----

FERROMAGNETIC RESONANCE LINE WIDTH: Many devices of interest at microwave and millimeter wave frequencies operate with the ferromagnetic material at resonance. At resonance a finite body in a metallic cavity or wave guide will absorb power at some well-defined frequency. Ferromagnetism arises due to the alignment and coupling of unpaired electron spins in matter. If a ferrite is placed in a d-c magnetic field, a torque is applied to the spinning electron, causing it to precess. The physical situation is depicted in Figure 6a. The magnetization, M, of the spinning electron precesses about the d-c field. Damping effects cause the precession angle to decrease until precession vanishes. This relaxation may occur in 10^{-10} seconds. If an a-c field is applied perpendicular to the d-c field, close to or at the resonance frequency, power will be absorbed in maintaining the magnetization of the electron precessing, Figure 6b. The a-c frequency

is at microwave and millimeter wave frequencies for hexagonal ferrites. Resonance, although a function of both d-c field and frequency of the a-c field, is usually studied, for experimental convenience, by holding the frequency fixed and varying the d-c field. Thus we plot absorption of a-c power versus applied d-c field at a fixed frequency. Line width is the width of the absorption peak at half-height and given in oersteds.

RESONANCE IN UNIAXIAL HEXAGONAL FERRITES: The equation for resonance of hexagonal ferrites with an easy axis (uniaxial) of magnetization is

$$\omega = \gamma(H_o+H_a) \qquad \text{Equation 1}$$

where, ω = resonance frequency in cycles/sec
γ = gyromagnetic ratio (a constant for a particular ferrite in mc/sec/oersted
H_o = applied magnetic d-c field in oersted
H_a = anisotropy field (a constant for a particular ferrite in oersteds)

The same equation is applicable to cubic ferrites. For the cubic compounds H_a is usually not more than several hundred oersteds. H_a ranges have been ten and twenty thousand oersteds in uniaxial ferrites. Thus Equation 1 shows that the range in frequency of uniaxial ferrites extends several orders of magnitude farther than that of cubic ferrites.

BARIUM FERRITE: Barium ferrite has an H_a = 17,000 oersteds. Taking γ = 2.8 and H_o = 7600 one finds ω = 55.2 kmc/sec. The line width of single crystals of barium ferrite in the form of spheres has been measured in this Laboratory (12) and by DuPre, et al, (13). At a frequency of 58 kmc/sec a line width of 53 oersteds has been observed. The single-crystal line width is compared with that for an oriented polycrystalline sample in Figure 7. Crystals containing small amounts of Al^{3+} for Fe^{3+} yielded line widths of 60 oersteds at 70 kmc/sec (12, 13). The presence of Al^{3+} had two desirable affects. It lowered the required saturation magnetization and increased the resonance frequency to 70 kmc/sec. When crystals with one atom of aluminum replacing one atom of iron, on a formula basis, become available, a resonance frequency of 120 kmc/sec will be reached.

PLANAR FERRITES: The equation for resonance in the case of the planar ferrites is

$$\omega = \gamma \quad H_o(H_o+H_a) \qquad \text{Equation 2}$$

The symbols have the same meaning as in Equation 1. Although some of these compounds have very large anisotropy fields (H_a), inspection of Equation 2 will show that the resonance frequency is much lower than for the axial type due to the square root. Resonance frequencies of uniaxial and several of these compounds are given in Table IV.

TABLE IV
Ferromagnetic Resonance Properties
of Hexagonal Ferrites

Compound	Symbol	Frequency of Line width* Measurement	Resonance Line width in oersteds: Single Crystal	Resonance Line width in oersteds: Poly-Crystal	Anisotropy Field H_a in Oersteds
$BaFe_{12}O_{19}$	M	58	53	1200	17,000
$BaFe_{11.5}Al_{15}O_{19}$	M	58	60	1200	19,700
$Ba_2Zn_2Fe_{12}O_{22}$	(Zn)Y	9.5	16	600	9,900
$BaCo_2Fe_{16}O_{27}$	(Co)W	15.0	2300	4200	------
$Ba_2Co_2Fe_{12}O_{22}$	(Co)Y	9.5	950	<4000	28,000
$Ba_4Co_2Fe_{36}O$	(Co)M_2Y		----	----	15,300
$BaZn_2Fe_{16}O_{27}$	(Zn)W	35.7	350	----	9,800

*in kmc/sec

Crystals of (Zn)Y have been studied in this Laboratory and industrial laboratories. Line widths as low as 16 oersteds (Table IV) have been observed at frequencies between 9 and 20 kmc/sec (14,15).

POWER LIMITER: S. Dixon (14) of this Laboratory has found, studying crystals of (Zn)Y with line widths between 60 and 150 oersteds, a large shift in the applied d-c field for resonance as microwave power is increased. This non-reciprocal behavior was observed between 9.5 and 15 kmc/sec. When 2000 watts of peak power was applied to a cavity containing a single crystal sphere, a shift of 600 oersteds was observed. The application of a gyromagnetic power limiter in this frequency range appears quite promising. Such devices are used to protect diodes and other components subject to burn-out with sudden power surges. In practice such a device would contain a sphere of (Zn)Y biased to resonance. The device couples power to another system.

SWITCHES: The principle just described may also be used as a switch in a microwave system. In addition if a planar ferrite is biased to resonance, and is coupling power at resonance, the removal of the biasing field would turn off the device. This biasing field can be quite small (16).

HARMONIC GENERATION (Frequency Doubling): Ferrites may be used to generate higher frequencies by employing the harmonics of the fundamental resonance frequencies. Frequencies not now attainable by tubes may thus be generated. The most important application is generation of higher frequencies from a fixed frequency microwave generator. Frequency doubling arises when magnetic asymmetry is present in the plane perpendicular to the precessing magnetization at resonance. If an asymmetry is introduced into the perpendicular plane, the magnetization precesses in an elliptical orbit. The longitudinal component of the magnetization will vary at double the frequency of the a-c field. The greater the magnetic anisotropy, the larger is

the ellipticity of the orbit and the better is the conversion efficiency. Asymmetry may arise either from demagnetization factors or magnetic anisotropy fields. Isotropic ferrites may be used for harmonic generation if they are properly oriented with the largest demagnetization field in the perpendicular plane. Hexagonal ferrites provide large anisotropy fields orders-of-magnitude larger than those present in cubic ferrites. Here the plane containing the easy axis or plane is oriented parallel to the biasing field.

Williams and Smith (7) have measured the second harmonic power output of several different ferrites. A maximum of 100 watts output at 10% efficiency was obtained with single crystals of barium ferrite. At the same input frequency of 35 kmc/sec cubic ferrites yielded a maximum of 8 watts. Planar ferrites should perform equally at the same and lower frequencies.

CONCLUSION: The investigations described in this paper demonstrate that large, pure, structurally sound, single-crystal specimens of the complex hexagonal ferrites can readily be grown. The method is further adaptable to the growth of crystals with various chemical substitutions. Such crystals have demonstrated new and enhanced electromagnetic properties.

The measurements of intrinsic properties on single crystals have resulted in a better understanding of the fundamental mechanisms involved in magnetic processes. The unique magnetic anisotropies of the planar and axial hexagonal ferrites have practical application at microwave and millimeter wave frequencies. The intrinsic narrow ferromagnetic resonance line widths have directed attention to special linear and non-reciprocal applications.

Single crystals of the hexagonal ferrites operate quite satisfactorily in devices at 50 and 70 kmc/sec (millimeter wave spectrum), permitting extension of communications to higher frequencies. A still further extension can be obtained by using these crystals for harmonic generation. Other hexagonal ferrite crystals, whose intrinsic properties are just now being studied, give promise of additional unique applications.

ACKNOWLEDGMENT: The authors are grateful to Messrs. I. Bady, S. Dixon, and G. McCall of this Laboratory and C. Buffler, Bomarc Laboratories, for many helpful discussions and for permission to use some of their microwave data before publication. We thank J. Mellichamp and R. Buder for spectrochemical analysis of many crystals. Dr. J. Kohn advised on the crystallography and reviewed portions of the manuscript.

LITERATURE CITATIONS

1. Jonker, G. H., Wijn, H. P. and Braun, P. B., Proc. of I.E.E. 104 249-254 (1956).

2. Adelsköld, V. Arkiv. Kemi. Minerali Geoli, 12A 29 (1938).

3. Went, J. J., Rathenau, G. W., Gorter, E. W., and Oosterhout van, G. W., Philips Tech. Rev. 13, 194 (1951).

4. Wijn, H. P. J., Nature, 170 707-708 (1952).

5. Kooy, C., Philips Tech. Rev. 19 286-89 (1958).

6. Brixner, L. H., J. Am. Chem. Soc. 80 4424 (1958).

7. Mones, A. H. and Banks, E., J. Phys. and Chem. Solids 4 217-22 (1958).

8. Gambino, R. J. and Leonhard, F., J. Amer. Ceram. Soc. 44 221-224 (1961).

9. DuPre, F. K., DeBitetto, D. J. and Brockman, F. G., J. Appl. Phys. 29 (1958).

10. Knick, P. and Kohlmeyer, E. J., Z. Anorg. u. allgem. Chem. 244 67-84 (1960).

11. Smit, J. and Wyn H. P. J., Ferrites John Wiley and Sons. New York pp 177-211 (1959).

12. Bady, I. private communication U.S.A.S.R.D.L.

13. DuPre, F. K. DeBitetto, Brockman, F. G., and Steneck, G. W. "Hexagonal Magnetic Materials for Microwave Applications" Final Report Signal Corp Contract DA36-039 SC-78071 Philips Labs. Irvington-on-Hudson N. Y. June 1958-1959.

14. S. Dixon - Private Communication U.S.A.S.R.D.L.

15. C. Buffler - Private Communication BOMARC Labs.

16. Bady, I. Trans. I.R.E. MTT-9 52 (1961).

17. Williams, G. W. and Smith, A. W. Electronics 35 58 (1962).

Figure 1. Anisotropic Hexagonal Ferrite Crystals

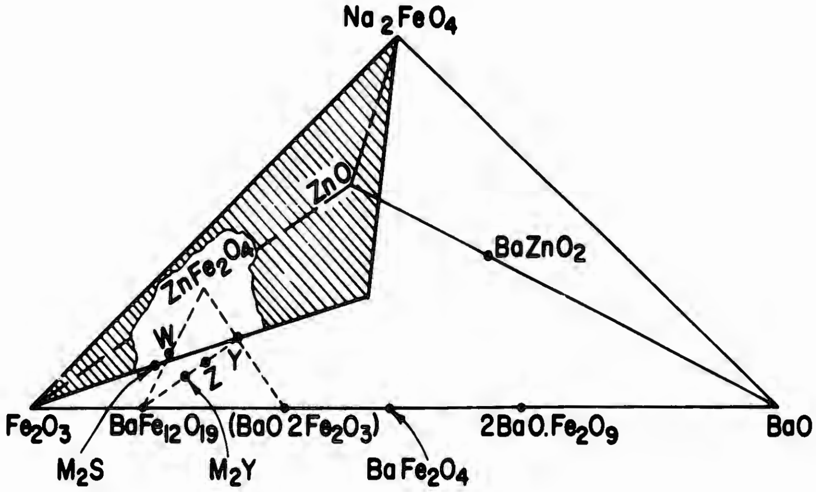


Figure 2. Composition Tetrahedron for the System
Na_2FeO_4-Fe_2O_3-BaO-ZnO

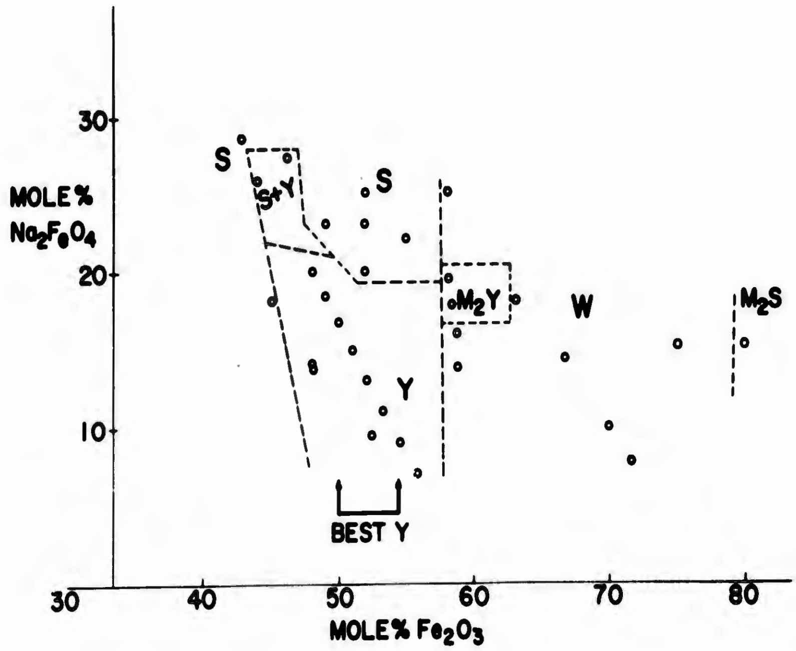


Figure 3. Crystallization in the BaO:ZnO 1:1 Composition Plane

Figure 4. (Zn)Y Crystals. Scale in inches

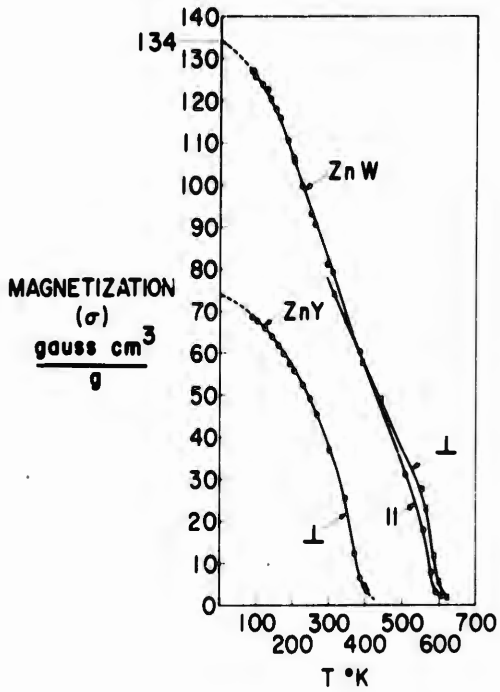


Figure 5. Magnetization versus Temperature for Single Crystals of (Zn)Y and (Zn)W
Applied field 10,500 oersteds
Field applied parrallel to c axis of crystal
Field applied perpendicular to c axis of crystal

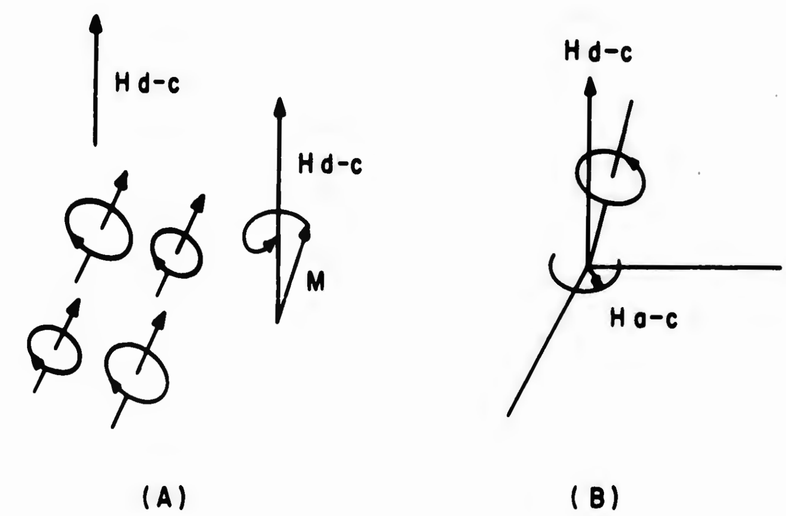


Figure 6(a) Precession in a d-c field
Figure 6(b) Resonance, d-c ⊥ a-c

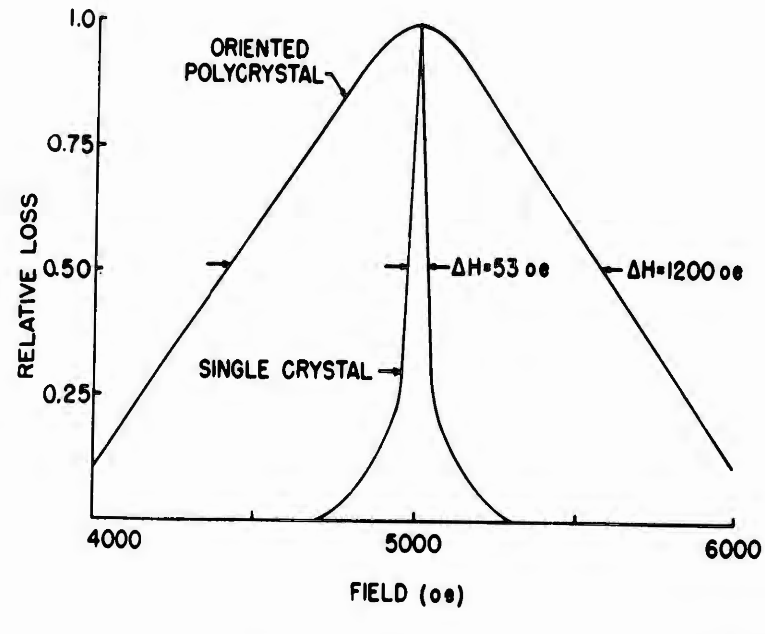


Figure 7. Ferromagnetic Resonance Line width in Barium Ferrite

MICROWAVE DIAGNOSTICS IN THE SHOCK TUBE

F. L. TEVELOW and H. D. CURCHACK
DEPARTMENT OF THE ARMY
DIAMOND ORDNANCE FUZE LABORATORIES
WASHINGTON 25, D. C.

INTRODUCTION:

The electromagnetic properties of an ionized gas or plasma may be determined uniquely by measurement of the change in phase and amplitude of a signal propagating through the gas.(1) A basic circuit for these measurements is the microwave bridge or interferometer.(2)

If a steady-state plasma exists, as in gas discharge tubes, flames, and rocket exhausts, propagation measurements can be made by manually nulling an attenuator and phase shifter in the reference arm of the interferometer.(3)(4)

For transient plasmas, Wharton and Slager made measurements at conditions where: (1) the effect of electron collisions is negligible (low attenuation) and (2) the transverse dimensions of the plasma are large compared with the incident signal wavelength.(5)

In the DOFL microwave propagation experiment, time-resolved plasma parameters are measured in a shock tube utilizing a microwave bridge circuit. It is assumed that a uniform, homogeneous slab of plasma completely fills a section of waveguide. The effects of electron collisions, reflections at boundary surfaces, and changing phase may then be rigorously included in the theoretical derivation. (6)(7) The condition of a closed waveguide is approximated by using a plasma thickness which is a fraction of the test signal wavelength.

Another current DOFL microwave experiment in the shock tube is the measurement of the noise power radiated by the plasma as it passes through the waveguide. For the conditions of the experiment, the radiation temperature may be related to the electron temperature. The test section again approximates a uniform slab enclosed in waveguide and as such, the noise power can be expressed rigorously in

terms of black body radiation, power reflection coefficient, and attenuation coefficient for the plasma at a particular frequency and temperature.

THEORY(DIELECTRIC-SLAB APPROXIMATION) (6)(7):

The theory for the propagation experiment assumes a uniform, homogeneous plasma slab completely enclosed by waveguide. A cross-sectional view of the plasma in the waveguide is shown in Fig. 1. Region 2, of thickness L, is the plasma; regions 1 are air-filled waveguide. A signal, E_i, is shown incident on the plasma at boundary A. The plasma impedance will determine the fraction of the incident signal that will pass into the plasma and the degree of attenuation and phase shift of the signal while traversing the plasma. E_r is the total signal reflected at A and E_t is the total signal transmitted at B. The amplitude and phase of the transmitted signal is uniquely determined by the plasma which can be described by the complex permittivity constant, ϵ_r, or by the complex propagation constant, γ, where

$$\epsilon_r = \epsilon'_r - j\epsilon''_r \tag{1}$$

and

$$\gamma = \alpha + j\beta \tag{2}$$

ϵ'_r is the real part and ϵ''_r the imaginary part of ϵ_r. α is the attenuation constant and β is the phase velocity constant. A complex transmission coefficient for the slab is defined as

$$\tau = E_t/E_i \tag{3}$$

Analysis of the wave contributions at surfaces A and B in Fig. 1 and of attenuation and phase shift in region 2 leads to

$$1/\tau = 1/2 \left[(\gamma_\epsilon/j\beta_g) + (j\beta_g/\gamma_\epsilon)\right] \sinh \gamma_\epsilon L + \cosh \gamma_\epsilon L \tag{4}$$

The subscript ϵ refers to constants for plasma-filled waveguide, subscript g refers to the constant for air-filled waveguide.

The angle and magnitude of τ are determined experimentally; α and β are then computed from equation (4). ϵ'_r and ϵ''_r can be determined from the following relations

$$1 - \epsilon'_r = (\lambda_o/\lambda_g)^2 \left[1 + (\alpha_\epsilon/\beta_g)^2 - (\beta_\epsilon/\beta_g)^2\right] \tag{5}$$

$$\epsilon''_r = 2(\lambda_o/\lambda_g)^2(\alpha_\epsilon/\beta_g)\ (\beta_\epsilon/\beta_g) \tag{6}$$

where λ_o is the free space wavelength and λ_g is the guide wavelength when filled with un-ionized gas. The plasma frequency, ω_p, and electron collision frequency, ν, can then be calculated from

$$\omega_p = \omega \left[(1 - \epsilon_r') + (\epsilon_r'')^2/(1 - \epsilon_r')\right]^{1/2} \tag{7}$$

$$\nu = \omega\epsilon_r''/(1 - \epsilon_r') \tag{8}$$

where ω is the incident signal frequency. The electron density is computed from

$$N = m\epsilon_o\omega_p^2 e^{-2} \tag{9}$$

where m is the electron mass, e, is the electron charge, and ϵ_o is the permittivity of free space. Equations 7, 8, and 9 are derived from the simple Lorentz force equation.

MICROWAVE TEST SECTION:

A schematic view of the microwave test section in the shock tube is shown in Fig. 2. The primary purpose of the test section is to pass the plasma through the waveguide without disturbing the flow or the plasma properties. The design is basically the same as that used previously by Lin.(9) A rectangular duct, with sharp leading edges, sections a portion of the plasma, which then flows through narrow slots in the waveguide. The remainder of the flow is directed into an expansion section. There is some loss to the walls, as in boundary layer growth, but these effects are negligible.(9) The windows, in conjunction with the internal walls, constrain the flow to the original streamlines. They are tuned to the test frequency and pass the microwave signal with less than 1% reflection. Except for the narrow slots in the sidewalls, (less than $\lambda/5$ wide at 10 kmc) the waveguide walls in the test section are continuous.

A photograph of the microwave test bench is shown in Fig. 3. The microwave test section of the shock tube can be seen on the right.

MICROWAVE CIRCUITRY:

A block diagram of the microwave circuitry is shown in Fig. 4. Low power x-band radiation is used (10 kmc, 20 mw) as the test signal and provision is made for monitoring both the frequency and power. This signal divides between the test branch and the reference branch. Various devices are used for equalizing signal levels and adjusting the phase between the two branches. The initial phase difference is kept at 90 degrees since this makes calibration easier. The isolators prevent reflections from feeding back and affecting power division and power levels. The test

signal and reference signal are fed into the Magic T where mixing occurs. Two outputs are obtained from the crystal diodes in the output arms of the Magic T. One signal level E_Σ is proportional to the vector sum of the two imput signals; the other E_Δ is proportional to the vector difference. These signals are displayed on an oscilloscope which has previously been triggered from a pickup located one foot up stream of the microwave section. Voltage outputs are obtained from the oscillogram deflections by comparison with deflection curves obtained using the precision attenuator. Calibration curves are obtained periodically while running experimental shots; but over a period of several months, these crystal output curves have not changed more than a few percent. The crystal diode attenuator is a dynamic calibration device, checked against the precision attenuator, and used as reference for each test shot. τ is determined from data computed from the simultaneous deflections of the two output crystals. This is given by (6) (7)

$$|\tau| = \left[(|E_\Sigma|^2 + |E_\Delta|^2 - 2|E_R|^2)/2|E_R|^2\right]^{1/2} \qquad (10)$$

$$\measuredangle = \sin^{-1}\left[(|E_\Sigma|^2 - |E_\Delta|^2)/4|\tau|\ |E_R|^2\right] + \beta_g L \qquad (11)$$

$|E_R|$ is the reference signal amplitude. Fig. 5 displays an oscillogram for a Mach 8 shock in argon. Time sweeps from left to right with blanking marks at 10 μsec intervals. The dynamic calibration mark indicating -3db with respect to initial power levels can be seen at 0.2 cm. For the first 3 cm, the test section is filled with undisturbed gas. The change in signal level at about 3.5 cm marks the arrival of the shock front. Conditions continue to change gradually for approximately 250 μ secs and then there is an abrupt deflection which marks the arrival of the cold driver gas. The upper curve is the output of the Magic T difference arm; the lower signal is the sum arm output.

THE SHOCK TUBE AND THE TEST PLASMA:

The shock tube has a 2 in. x 2 in. cross section with the microwave test section located approximately 15 ft. downstream from a scored aluminum diaphragm. This diaphragm is designed to burst at a pre-determined pressure and initially separates the high-and low-pressure sections of the shock tube. In operation, the low-pressure section is evacuated and the test gas is introduced at the desired pressure. The driver-gas pressure is then increased, bursting the diaphragm. For a given test gas and initial test pressure, the shock Mach number is controlled by varying one or more of the following parameters: (1) the depth of score of the diaphragm (2) type of driver gas, and (3) the temperature of the driver gas.

When the diaphragm bursts, a shock wave propagates into the low pressure gas. There is a sudden increase in the translational energy as the molecules pass through the shock front and get carried along down the shock tube. Behind the shock, there is a randomization of this energy into various other modes such as rotation, vibration, dissociation, excitation, and ionization, depending on the species of gas involved. The rates at which the energy is distributed amongst the modes varies but the first three rates are usually fast compared to the excitation and ionization rates. Given sufficient initial energy, ionization will occur and with sufficient test time equilibrium will be observed. Equilibrium values behind the shock can be computed from the initial equilibrium conditions by using the Rankine-Hugoniot, or shock tube relations.

The test gases for which results will be presented are argon and dry air. In argon, the initial translational energy goes into excitation and ionization of the atom. The main electron producing reactions (10) are:

1. $A + A^* \rightarrow A^+ + A + e$
2. $A + A \rightarrow A^+ + A + e$
3. $A + e \rightarrow A^+ + 2\,e$

where A is argon, A* represents an excited state, and A^+ is the ionized argon atom.

Ionization levels are determined initially by reaction rates for processes 1 and 2 . When the ionization level reaches approximately 10% of the equilibrium level, the rate of ionization is dominated by process 3 and rises rapidly to equilibrium. These rates are such that equilibrium conditions could not be seen in this shock tube for shock Mach numbers less than 10. (11) If process 3 has not commenced, it is reasonable to assume that the electron temperature, T_e, obeys a Maxwellian velocity distribution and is equal to the plasma temperature, T_g. If process 3 is dominant, then T_e is slightly lower than T_g because each time reaction 3 occurs, an amount of kinetic energy equal to the ionization energy of argon is lost by the electron gas. However, for an ionization degree, x, greater than 10^{-3}, sufficient electron-electron collisions occur so that the distribution is again Maxwellian and a temperature T_e can be ascribed to the electron gas.(10)

Air being a diatomic gas permits randomization of the translational energy amongst all of the modes previously mentioned. With sufficient initial translational energy, ionization will occur as given by the following processes:

1. $O + M \rightarrow O^{+} + e + M$
2. $N + M \rightarrow N^{+} + e + M$
3. $N + NO \rightarrow O^{+} + N_2 + e$
4. $N + O \rightarrow NO^{+} + e$

where M is any atom or molecule present. Process 4 is the dominant ion producing reaction because of its relatively low ionization energy. As explained previously, the electron gas may be described by a temperature, T_e, equal to the gas-temperature.

MICROWAVE RADIATION:

A block diagram of the microwave circuitry used in the noise radiation experiment is shown in Fig. 6. The same test section used for the propagation measurements is utilized for this experiment. Although shown enlarged, the plasma thickness is narrow compared to its height and the waveguide windows constrain the plasma as before.

The output of the transient plasma is fed into the microwave mixer, amplified, envelope detected, and then fed to a pre-triggered oscilloscope where it is photographically recorded. System calibration is accomplished by means of a temperature standard (calibrated fluorescent tube mounted in waveguide) and a precision attenuator. Preliminary measurements obtained at DOFL using this system have been reported earlier. (13)

The radiation in the microwave region is due to bremsstrahlung by the free electrons of the plasma. The radiation temperature is defined by Kirchoff's law

$$P_\omega = B(\omega, T)\, A_\omega \tag{12}$$

where P_ω is the radiation intensity in the frequency interval between ω and $\omega + d\omega$, $B(\omega, T)$ is the blackbody radiation intensity and A_ω is the absorptivity of the plasma. When the electrons are in thermodynamic equilibrium with the radiation and when the electrons have a Maxwellian distribution of velocities, the radiation temperature is the same as the electron temperature. (14)

For a dielectric slab enclosed by waveguide (uniform temperature and a constant absorption coefficient), the observed power is given by

$$P = (1 - \Gamma)\,(1 - \Gamma \exp -\alpha L)^{-1}\, B(\omega, T)\,(1 - \exp -\alpha L) \tag{13}$$

where Γ is the power reflection coefficient, α is the attenuation coefficient, and L is the thickness of the plasma slab. Thus, to determine the radiation temperature quantitatively, time functions of Γ and α must be determined.

Figure 7. is an oscillogram from reference 13 showing the radiation at 9300 mc from a transient argon plasma. Time sweeps from left to right at 50 μ secs/cm. Initially, the receiver sees a load at room temperature and the output is a d.c. level with fluctuations. As the plasma passes through the section, the average d.c. level and the fluctuations increase. The average deflection for this shot assuming $\alpha = 0$ corresponded to a temperature of 4000°k. Equilibrium theory predicted a gas temperature of 6000°k. Corrections for reflection and absorption losses would have increased the computed value for the radiation temperature but plasma parameters were not determined for this experiment. Simultaneous propagation and noise measurements have since been made. Plasma parameters and noise characteristics were obtained for a Mach 8 shock in argon. These measurements were made at 10 kmc to demonstrate feasibility only. The measurements will be repeated at 9300 mc where the receiver noise figure is much better. The major problem, as can be seen from Fig. 7, will be separation of the average d.c. deflection from the random fluctuations.

RESULTS:

Time variation of electron density at initial pressures of 1, 2 and 5 mm Hg. are presented in Fig. 8 for a Mach 10.0 shock in air. The abscissa is the time that a particle has been behind the shock and is related to observed time by a factor of the density ratio across the shock. Zero time is recorded as the point at which an initial signal deflection is first noted and may be several μ secs in error. The rate of rise for the first 100 μ secs was too rapid for resolution and is not shown. It is seen that peak values of electron density and the length of time that the density remains at this steady-state value both increase with an increase in pressure. From the shapes of the curve, it can be assumed that the peak value at 1 mm is essentially also the steady-state value.

Peak values of electron density (steady-state) at initial pressures of 1,2, and 5 mm Hg are given in Figure 9. as a function of Mach number. Electron density is seen to increase with pressure and Mach number. The lines represent computed values of equilibrium electron density using conditions behind the shock from Hochstim (16) and equilibrium constants for argon-free air (17). Measured values are approximately a factor of 2 greater than equilibrium values for all values. These results are similar to that obtained by Lin at high Mach numbers and lower pressure using a magnetic induction probe (18).

Time variation of electron density for shocks in argon and in air are shown in Fig. 10. The data are presented primarily to show the marked difference in the rates and characteristic curves. It can be seen that the electron density for argon continually increases and does not reach equilibrium in the available test time.

Real time for air is approximately 1/10 that of particle time while real time for argon is approximately 1/4.

Collision frequency for electrons in air is given at initial pressures of 1,2, and 5 mm Hg as a function of Mach number, Fig.11. The symbols represent maximum values of collision frequency observed during the test shot while the lower extreme of the line extending from the symbols gives the value of collision frequency measured at the same instant of time as the steady-state equilibrium electron density. The theoretical lines were extrapolated from curves presented by Bachynski (19) for temperatures between 3000 and 4000°k. Values of collision frequency were calculated by Bachynski assuming the mean free path of the electrons independent of electron velocity and using the relation

$$\nu = \bar{c}\Sigma n_j Q_j \qquad (14)$$

where n_j is the number density of species j, Q_j is the Maxwell-averaged total electron collision cross section of the species and

$$3\bar{c} = 4\ (8kT/\pi m)^{1/2}, \qquad (15)$$

k being Boltzmann's constant. Collision cross sections for the neutral species were obtained from available data.

Fig. 12 gives measured values of peak attenuation as a function of Mach number and pressure. It is seen that attenuation increases with an increase in either pressure or Mach number. Values as high as 20 db/cm were observed.

DISCUSSION:

The microwave system, including the slotted waveguide section and half-wavelength windows, was calibrated originally by dropping bars of teflon, lucite, and brass through the test section. Measured values of the dielectric constants were compared to handbook data and values determined by the use of slotted line techniques. All results agreed to within 1%. The brass bar demonstrated that the electromagnetic field was confined essentially to the dimensions of the waveguide. The effect of a fringe field was not discernible.

Precision of numerical data obtained from the oscillograms varies with the amount of signal deflection. Precision of the measurements averages ±10% for electron density and ±20% for collision frequency. The assumptions of a uniform test gas through the thickness and the approximation to a closed waveguide have been substantiated. Since conditions change with time, the plasma is not uniform along the length and an average value over one inch is being determined. Increased space resolution will be obtained when a theoretical analysis of a space varying dielectric constant is completed.

BIBLIOGRAPHY:

(1) Goldstein, L., "Electrical Discharge in Gases and Modern Electronics," Advances in Electronics, 7 Academic Press, N.Y., N.Y. (1955),431-5.

(2) Whitmer, R.D., "Microwave Studies of the Electron Loss Process in Gas Discharge," Phys. Rev., 104 (1956), 572-5.

(3) Goldstein, L., Lampert, M.A., and Geiger, R.H., "Determination of Electron Density and Collision Frequency in a Gaseous Discharge by Microwave Propagation," Elec. Comm., 29(1952)243-5.

(4) Shuler, K.E., and Weber, J., "Microwave Investigation of the Ionization of Hydrogen-Oxygen and Acetylene-Oxygen Flames," J. Chem. Phys., 22 (1954) 491-502.

(5) Wharton, C.B., and Slager, D.M., "Microwave Determination of Plasma Density Profiles," J. Appl. Phys., 31 (1960) 428-430.

(6) Tischer, F.J., "Measurement of the Wave Propagation Properties of Plasma in the Microwave Region," Ohio State Univ Rpt 941-1 (1957) 17.

(7) Tevelow, F.L., and Curchack, H.D., "Shock-Tube Microwave Propagation Measurements Using the Dielectric-Slab Approximation" DOFL TR-962 (1961)41.

(8) Bekefi, G., and Brown, S.C., "Emission of Radio Frequency Waves from Plasmas," Am. Jour. Phys., 29 (1961) 404-8.

(9) Lin, S.C., Ionization Phenomenon of Shock Waves in Oxygen-Nitrogen Mixtures," AVCO Res. Rpt. 33, (1958) 4, 10.

(10) Bond, John W., "Structure of a Shock Front in Argon," Phys. Rev., 105 (1957) 1683-93.

(11) Petschek, H., and Byron, S., "Approach to Equilibrium Ionization Behind Strong Shock Waves in Argon," Annals of Physiks, 1('57)270.

(12) Atallah, S., "The Kinetics of Air in a Hypersonic Shock Wave," AFCRL Rpt 761, (1961) 19.

(13) Peperone, S., "X-Band Measurement of Shock-Tube Plasma Temperature," J. Appl. Phys., (1962).

(14) Bekefi, G. and Brown, S.C., "Microwave Measurements of the Radiation Temperature of Plasmas," J. Appl. Phys., 32(1961) 25.

(15) Bekefi, G. and Brown, S.C., "Emission of Radio Frequency Waves from Plasmas," Am. J. of Phys., 29 (1961) 406.

(16) Hochstim, A.R., "Gas Properties Behind Shocks at Hypersonic Velocities, I. Normal Shocks in Air," Convair Rpt ZPH(GP)002,'57.

(17) Private Communication from J. Hilsenrath, National Bureau of Standards, Wash., D.C. (1961).

(18) Lin, S.C., Neal, R.A., and Fyfe, W.I., "Rate of Ionization Behind Shock Waves in Air, 1. Exp Results," AVCO Res Rpt 105,'60.

(19) Bachynski, M.P., Johnston, T.W., and Shkarofsky, I.P.,"Electromagnetic Properties of High-Temperature Air," Proc. of IRE, (1960) 347.

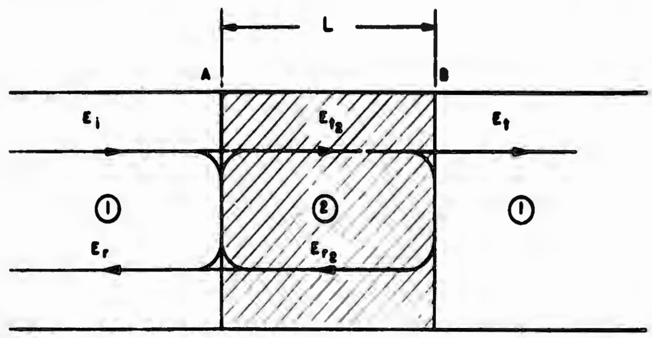


Figure 1. Voltage scattering-
schematic diagram

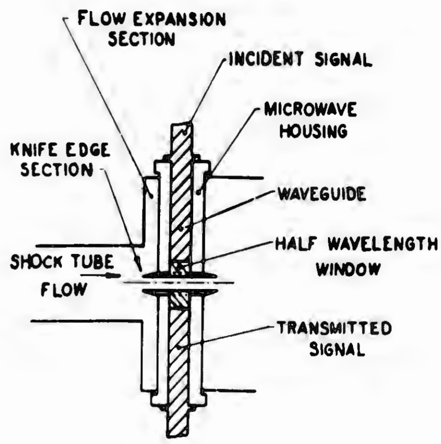


Figure 2. Shock-tube microwave test
section- schematic diagram

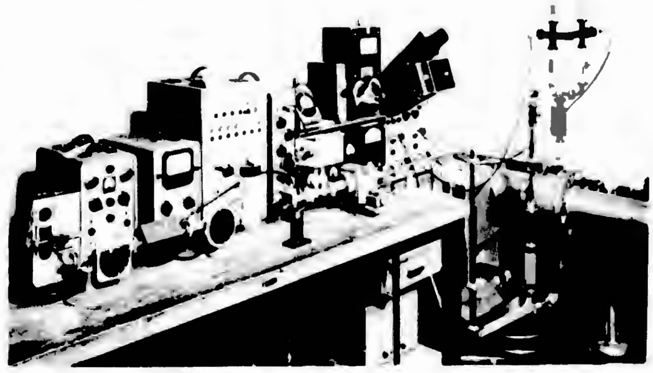


Figure 3. Microwave test bench and
shock-tube test section

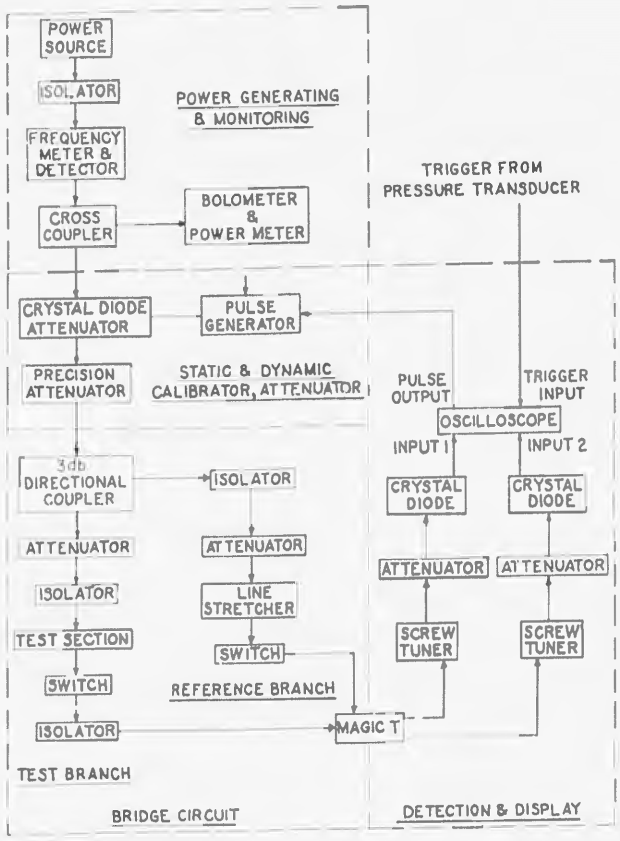


Figure 4. Microwave Circuitry-
block diagram

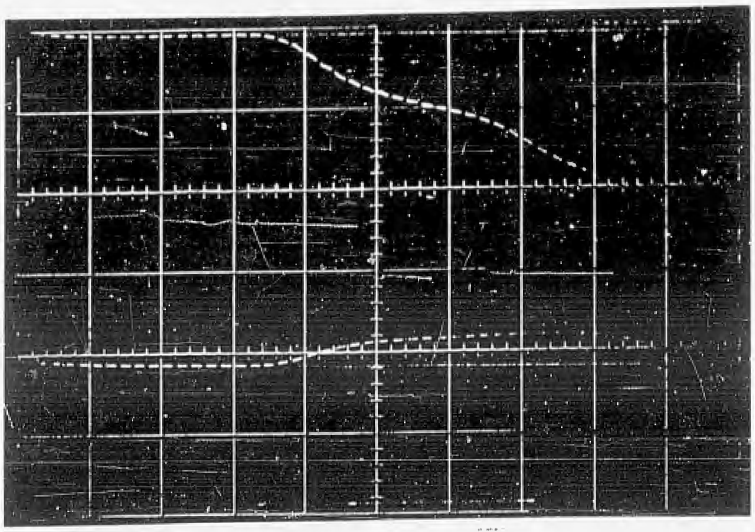


Figure 5. Magic T crystal diode
voltage output

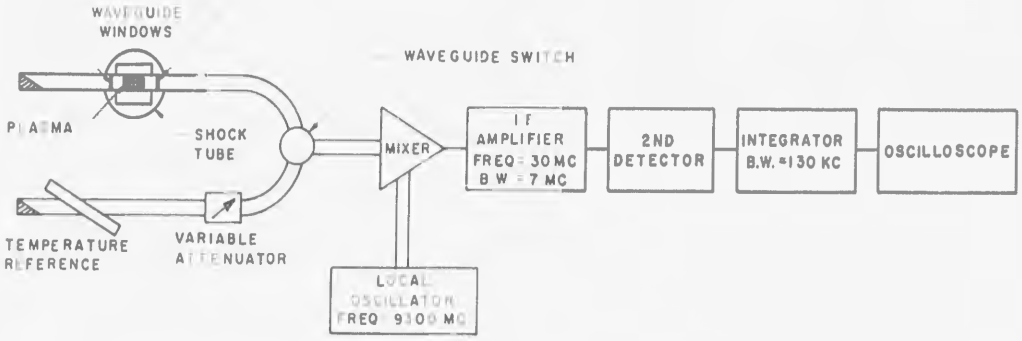


Figure 6. Microwave noise radiation circuitry-block diagram

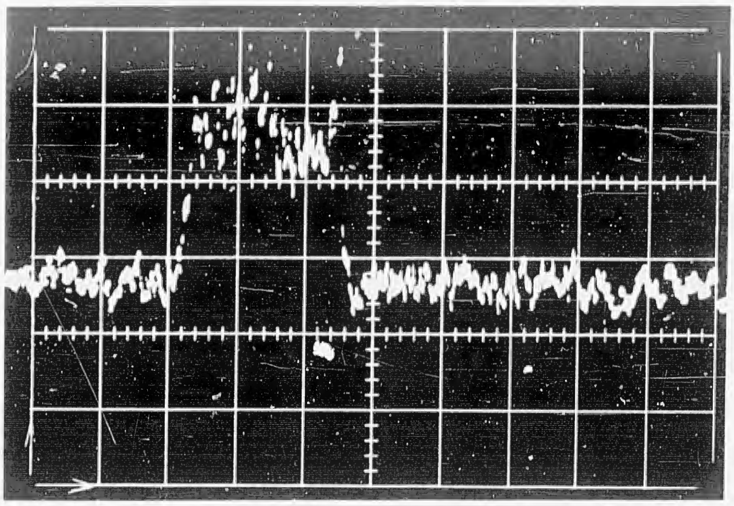


Figure 7. Receiver output as affected by the plasma radiation

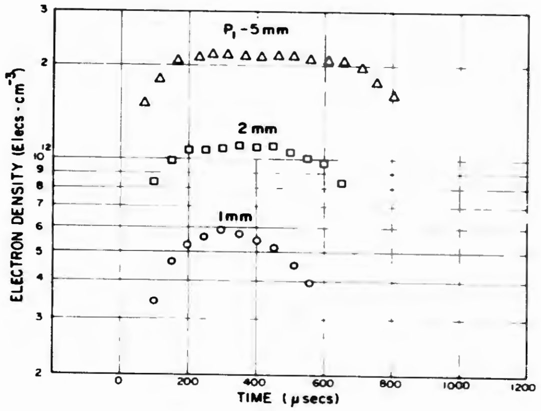


Figure 8. Electron density in air as a function of particle time behind the shock for initial pressures of 1, 2, and 5 mm Hg. at Mach 10 and an initial temperature 300° K

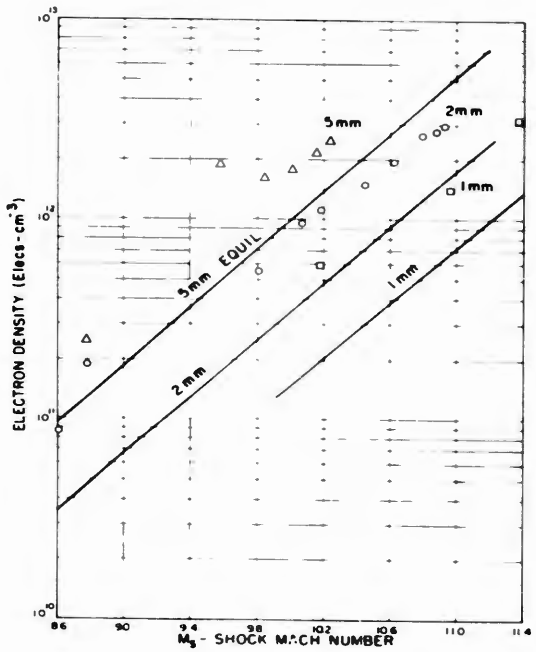


Figure 9. Peak electron density in air as a function of shock Mach number at initial pressures of 1, 2, and 5 mm Hg. and an initial temperature of 300° K. Lines represent equilibrium values of electron density behind the shock computed from equilibrium constants for argon-free air

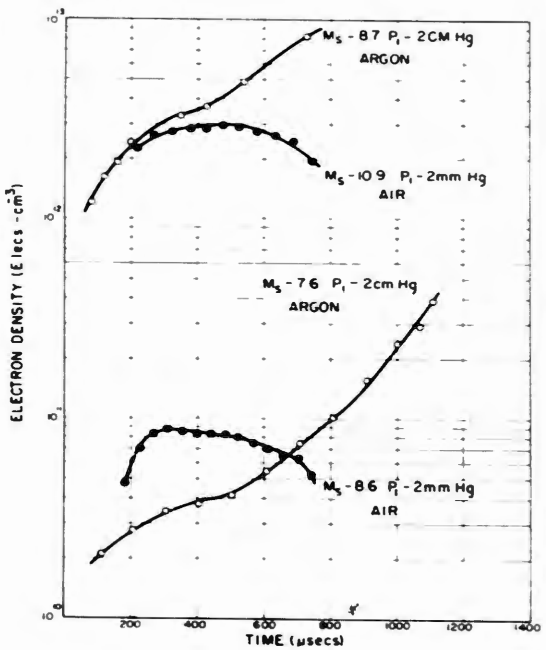


Figure 10. Electron density as a function of particle time behind the shock. Circles are measurements in argon at an initial pressure of 2 cm Hg at Mach 7.6 and 8.7. Triangles are measurements in air at an initial pressure of 2 mm Hg. at Mach 8.6 and 10.9 initial temperature in all instances 300° K

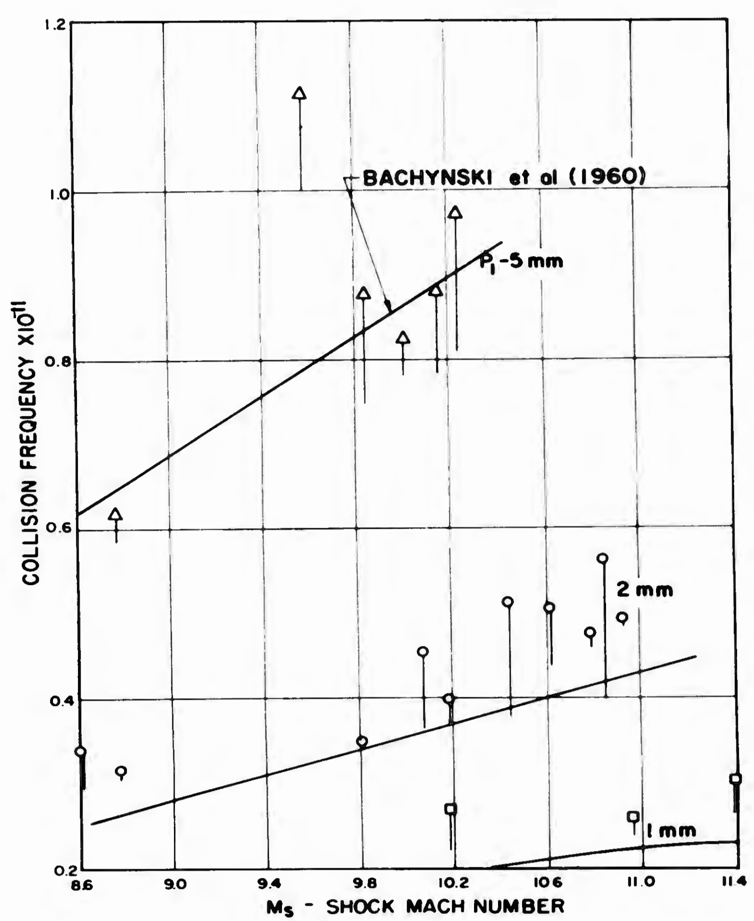


Figure 11. Electron collision frequency in air as a function of Mach number at initial pressures of 1, 2, and 5 mm Hg and initial temperatures 300° K

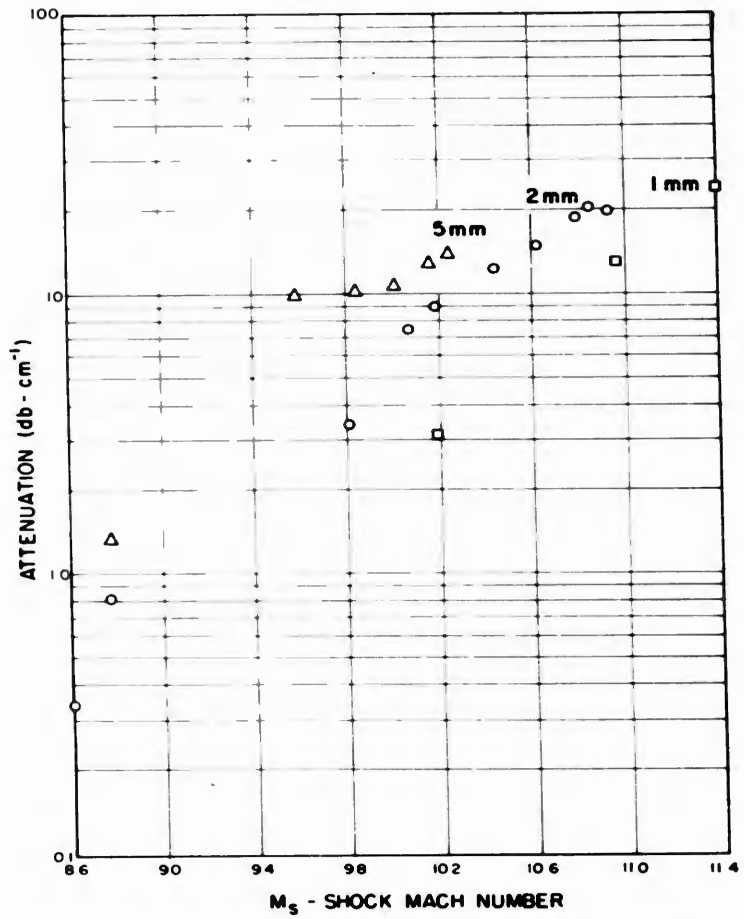


Figure 12. Peak electromagnetic attenuation in air at 10 KMC as a function of Mach number for initial pressures of 1, 2, and 5 mm Hg and initial temperature of 300° K

MESOSPHERIC DENSITY VARIABILITY BASED ON RECENT METEOROLOGICAL ROCKET MEASUREMENTS

OTTO W. THIELE
U. S. ARMY SIGNAL MISSILE SUPPORT AGENCY
WHITE SANDS MISSILE RANGE, NEW MEXICO

I. INTRODUCTION

Until recently, most of the techniques for determining high-altitude density, pressure, and temperature employed very expensive and complicated rocket systems. Obviously, such systems have not been feasible for investigating the high atmosphere with a sufficient quantity of soundings to permit a comprehensive variability study. However, with the advent of the Meteorological Rocket Network, in 1959, an increasing volume of data is becoming available for a more detailed investigation of the meteorological aspects of the region above balloon altitudes.

The primary system being used in the Meteorological Rocket Network for measurements other than wind consists of a single-channel telemetering unit (1,2) utilizing a 10-mil ceramic bead thermistor as the sensor (3,4). This instrumentation is carried aloft in an Arcas sounding rocket (5,6,7) and expelled on a parachute at apogee. The altitude range varies from approximately 225,000 feet for sea-level launches to approximately 250,000 feet for a launch elevation of 4000 feet such as at White Sands Missile Range.

The temperature is sensed during descent and telemetered to standard AN/GMD-1 (Ground Meteorological Detector) radiosonde equipment. The temperature data are recorded and reduced in much the same manner as balloon-borne radiosonde data and pressure and density are calculated using a "tie-on" pressure from a compatible balloon-borne radiosonde. The metalized parachute is tracked for position data (8) which, of course, provides wind data.

II. DISCUSSION OF DATA

This paper is concerned with data obtained from Arcas meteorological rockets fired at White Sands Missile Range (WSMR), New

Mexico, and Fort Churchill (FT CH), Canada, in connection with the Meteorological Rocket Network and R&D efforts. A reasonable sample of data was acquired during the seasons of summer 1960, fall of 1960, and winter of 1961 plus a small amount of interim data. A total of 51 temperature profiles were obtained during this period for the density derivations.

Emphasis is placed on seasonal and latitudinal variability. Unfortunately, however, there is no winter data available for inclusion at this time for the northern latitudes. "Quick-look" temperature data recently (January and February 1962) obtained from Fort Churchill, Canada, and Fort Greely, Alaska, indicate that the late fall data obtained at Fort Churchill (1960) are reasonably representative of northern latitude winter data.

Figures 1 and 2 for summer at WSMR and FT CH show that the maximum and minimum range at each station is small and that there is little latitudinal variation in the summer. The fall data, figures 3 and 4, also show that the maximum and minimum range during the particular season is small at each station; however, the latitudinal variation between the two stations begins to vary considerably. Figure 5 containing the WSMR winter data shows a somewhat larger maximum and minimum range during the particular season. As pointed out earlier, there is no northern latitude winter data available for direct comparison.

The mean density profiles in figure 6 show quite clearly the progressive decrease in density from summer to winter (or vice versa) at both stations. It is also evident from figure 6 that the seasonal variability is small at WSMR and quite large at FT CH. This consolidated plot again points out the near absence of latitudinal variability in the summer and the considerable increase in latitudinal variability in the fall and also implies possibly even more latitudinal variability in the winter.

Figure 7 presents the maximum and minimum as well as the mean of all the data. The overall variation begins to be quite large above 150,000 feet, varying between 40 and 50 percent up to 200,000 feet. The overall mean generally agrees quite well with the Air Research and Development Command (ARDC) standard atmosphere for density profile. However, the indication is that the atmosphere above 140,000 feet is slightly less dense than the ARDC model (9). It should be pointed out that in all the figures the extreme high end of the data is not representative because of the relatively few samples of data at those altitudes.

The small amount of interim data presented in figure 8 is generally consistent with what would be expected between seasons. The September data were taken during an "International Geophysical Interval." The Fort Churchill April and May (Spring 1960) data also

included in figure 8 represent the mean of two low-altitude shots obtained during that period.

Although primary emphasis is on density, the general features of the pressure variability are presented in figures 9 and 10 since pressure data are obtained in the density derivations. The seasonal and latitudinal variation of pressure follow essentially the same pattern as does the density variability and therefore will not be discussed in detail.

III. METHOD OF REDUCTION

With the reduced temperature data, density and pressure values are calculated using the hydrostatic equation and the equation of state. The computations can be made quickly for immediate application.

The hydrostatic equation, $dp = -\rho g dz$

is modified by substituting from the equation of state,

$$dp = -\frac{pmg}{RT'} dz$$

where T' is the virtual temperature. The equation is then integrated from the highest pressure measurement p_o at height z_o obtained from the balloon-borne radiosonde taken in conjunction with the rocket firing, to the height z, having the pressure p_1, as follows:

$$\int_{p_o}^{p_1} \frac{dp}{p} = \frac{-mg}{R} \int_{z_o}^{z_1} \frac{1}{T'} dz = \ln \frac{p_1}{p_o}$$

$$p_1 = p_o e^{\frac{-mg}{R}\int_{z_o}^{z_1} \frac{1}{T'} dz}$$

or as further solv . and reduced for computer use

$$p_1 = p_o e^{\frac{-g(z_1 - z_o)}{R' T'_m}}$$

where $R' = R/m$, the gas constant for dry air, and T'_m is the mean virtual temperature in absolute degrees Kelvin through the layer $(z_1 - z_o)$. As used here, T'_m is just the equally weighted mean

temperature through the layer, since variations in composition at these altitudes is considered negligible.

The heights are geometric and are known quite accurately from radar data. An arbitrary layer thickness of 5,000 feet is used for these calculations and g is the gravity value at the mean height of the layer.

The computation is reiterated, with each succeeding calculated pressure (p_1) and height (z_1) at the top of the layer becoming the new p_o and z_o, to the maximum height z_n of the temperature measurement.

Having computed the pressures at the top of each layer, the density at each of these levels is calculated using the equation of state for dry air,

$$\rho = \frac{P}{R'T}$$

where T is the measured temperature in absolute degrees Kelvin at the particular level.

The complete problem has been programmed for use with the Philco S-2000 digital computer. The input data consist of the initial pressure p_o (dynes per square centimeter), initial height z_o (feet MSL), maximum height z_n (feet MSL), t_m' (°C), values for each layer and t (°C) values at the level representing the top of each corresponding layer. Gravity values, the height increment (5,000 feet), the gas constant for dry air, and conversion factors are preprogrammed.

It should be pointed out here that while the density values represent a specific calculation at each level, some detail evidenced between levels is omitted. The method of calculation outlined above was used for convenience at the time. However, a system based on the significant levels of temperature is presently being programmed that will result in density information in as much detail as is available from the temperature data.

IV. ERROR ANALYSIS

Although the primary intent of this paper is to present and discuss some of the data derived from Arcas meteorological rockets, it necessarily follows that at least some attention should be given to possible sources of error, as well as the probable range of known errors.

Of primary importance with the method of calculation used here, is the availability of a reasonably accurate "tie-on" pressure obtained from an associated balloon-borne radiosonde measurement. The

time and space variability of such a supporting measurement is necessarily inherent and cannot be accurately resolved. Every effort is made to obtain a pressure measurement at 100,000 feet within 30 minutes of the firing. In support of the data contained in this paper, approximately one-third of the measurements at balloon-burst altitude were within 30 minutes, two-thirds within an hour, and all within two hours.

While the support radiosondes are released in the vicinity of the rocket launching area, the space variability at burst can be considerable. However, it could normally be expected to be within 50 miles with a probable maximum of 100 miles.

Several radiosonde measurements are made daily from different locations at White Sands Missile Range, and a close inspection of several sets of these data (a set is all of the measurements on a given day at the same altitude) reveals differences of only two- or three-tenths of a millibar in the vicinity of 100,000 feet.

Errors due to time and space variability, when using compatible radiosonde data, are assumed to be insignificant. The overall accuracies of radiosonde measurements are discussed at considerable length in references (10,11,12,13,14).

Whenever possible, the pressure measurement is obtained with a hypsometer type radiosonde (10) which is generally acknowledged to be the most accurate system now available for field use.

The pressure values are determined as closely as possible to a tenth of a millibar for application in the computations. A quick inspection of the data will show that in the region of 100,000 feet, a one-tenth millibar error is approximately one percent which, of course, produces a one-percent error in the density calculation.

Next, it is necessary to consider the only other significant source of error -- the temperature measurement. A theoretical study by Jehn, Wagner and others (4,15) of the thermistor and instrumentation used, indicates that the temperature can be measured to within one degree up to 57 kilometers. Above this altitude the data become somewhat uncertain and a satisfactory technique for correcting the data has not yet been determined. However, from comparative data available, it appears that the temperature measurements are accurate to within 5 degrees from 57 kilometers up to approximately 68 kilometers.

The effect of small temperature errors relative to the computed density is not large. A two-degree temperature error will result in a density error of approximately one percent.

Up to an altitude of 57 kilometers, the computed densities are considered to be at least as accurate as density data derived from balloon-borne radiosondes.

Although accuracies within two percent are probable for more than half of the density data presented, all data are considered accurate to within five percent.

Of general significance is the uniformity of a particular group of data, implying that any possible errors are at least systematic, and do not detract materially from the overall value of the results from a comparison standpoint.

V. SUMMARY AND CONCLUSIONS

By using the method previously described, density and pressure data can be obtained from 30 kilometers to at least 60 kilometers economically and on a timely basis for direct application to space programs, forecasting, etc. Accuracies within two percent are presently attainable under ideal circumstances. However, the data presented here are considered to be accurate within five percent up to an altitude of approximately 68 kilometers.

These data tend to verify many of the significant features of the atmosphere between 30 and 60 kilometers that have been evident from previous experimental data (16,17,18,19,20,21,22,23) such as the wider seasonal range at northern latitudes, the greater variation in winter at both stations, and the overall variations -- as much as fifty percent. The need for both seasonal and latitudinal standard atmospheres where standards are necessary is clearly demonstrated. For an overall standard, the total data indicate that both density and pressure are slightly lower, from 40 to 60 kilometers, than the ARDC Standards. Doubtlessly, the increasing volume of data becoming available through the Meteorological Rocket Network will more accurately define these parameters in the upper atmosphere.

REFERENCES

1. Clark, George Q., "Arcas Nose Cone Alpha," Schellenger Research Laboratories, Texas Western College, El Paso, Texas, October 1959.

2. Clark, George Q., "Development of a Rocket Telemetry Package for the Meteorological Rocket Network," paper presented at the Joint Instrument Society of America -- American Meteorological Society Meeting in September 1961.

3. Ballard, H. N., "Response Time of and Effects of Radiation on the VECO Bead Thermistor," Schellenger Research Laboratories, Texas Western College, El Paso, Texas, April 1961, Contract DA-29-040-ORD-2410.

4. Wagner, N. K., "Theoretical Time Constant and Radiation Error of a Rocketsonde Thermistor," Journal of Meteorology, Vol 18, No 5, October 1961.

5. Webb, W. L., K. R. Jenkins, and G. Q. Clark, "Flight Testing of the Arcas," Technical Memorandum 623, U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, May 1959.

6. Webb, W. L., and K. R. Jenkins, "Application of Meteorological Rocket Systems," Journal of Geophysical Research, Vol 64, No 11, pp 1855-1861, November 1959.

7. Jenkins, K. R., W. L. Webb, and G. Q. Clark, "Rocket Sounding of High Atmosphere Meteorological Parameters," IRE Transaction on Military Electronics, MIL-4, Nos 2 and 3, 238-243, 1960.

8. Beyers, N. J., and O. W. Thiele, "Meteorological Rocket Wind Sensors," Special Report 41, U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, August 1960.

9. Handbook of Geophysics (Revised Edition) United States Air Force, Air Research and Development Command, Air Force Research Division, Geophysics Research Division, 1960.

10. Conover, W. C., and W. G. Stroud, "A High Altitude Radiosonde Hypsometer," Journal of Meteorology, Vol 15, No 1, pp 63-68, February 1958.

11. Air Weather Service, "Accuracies of Radiosonde Data," Technical Report 105-133, Headquarters, Air Weather Service MATS, USAF, September 1955.

12. Leviton, R., "Height Errors in a Rawin System," Air Force Surveys in Geophysics, AFCRC, TN-54-26, No 60, Air Research and Development Command, December 1954.

13. Shaw, J. H., "Numerical Evaluation of Radiation Errors of Radiosonde Temperature Element ML 419 at Altitudes up to 150,000 feet," Technical Report 3, The Ohio State University Research Foundation, March 1959, Contract DA-36-039-SC-78153.

14. Cline, D. D., "Error Analysis of Carbon Reference Resistors for Radiosonde AN/AMT-4 ( )," Engineering Report E-1152, Signal Corps Engineering Laboratories, February 1955.

15. Jehn, K. H., N. K. Wagner, J. R. Gerhardt, D. R. Haragan, "Wind and Temperature in the Atmosphere between 30 and 80 Km," Third Quarterly Technical Report, The University of Texas Electrical Engineering Research Laboratory, March 1961, Contract DA-23-072-ORD-1564.

16. Quiroz, Roderick S., "Seasonal and Latitudinal Variations of Density in the Mesosphere (30 to 80 Kilometers)," Journal of Geophysical Research, Vol 66, No 7, July 1961.

17. Smith, Orvel E. and H. B. Chenoweth, "Range of Density Variability from Surface to 120 Km Altitude," National Aeronautics and Space Administration Technical Note D-612, July 1961.

18. Jones, L. M., J. W. Peterson, and others, "Upper Air Density and Temperature: Some Variations and an Abrupt Warming in the Mesosphere," Journal of Geophysical Research, Vol 64, No 12, December 1959.

19. Bandeen, W. R., W. G. Stroud, and others, "The Measurement of Temperature, Densities, Pressures and Winds Over Fort Churchill, Canada, by means of the Rocket Grenade Experiment," U. S. Army Signal Research and Development Laboratories, Fort Monmouth, New Jersey, USASRDL Technical Report 2076, November 1959.

20. Stroud, W. G., and others, "Rocket Grenade Measurements of Temperatures and Winds in the Mesosphere over Fort Churchill, Canada," Journal of Geophysical Research, 65(8):2307-2323, August 1960.

21. Jones, L. M., and others, "Upper-Air Densities and Temperatures from Eight IGY Rocket Flights by the Falling-Sphere Method," IGY World Data Center A, Rocket Report Series, No 5, December 1959.

22. Nordberg, W., and W. G. Stroud, "Results of IGY Rocket-Grenade Experiments to Measure Temperatures and Winds above the Island of Guam," Journal of Geophysical Research, Vol 66, No 2, pp 455-464, February 1961.

23. Spencer, N. W., R. L. Boggess, and D. Taeusch, "Pressure, Temperature and Density to 90 Km over Fort Churchill," IGY World Data Center A, Rocket Report Series, No 1, pp 80-90, July 1958.

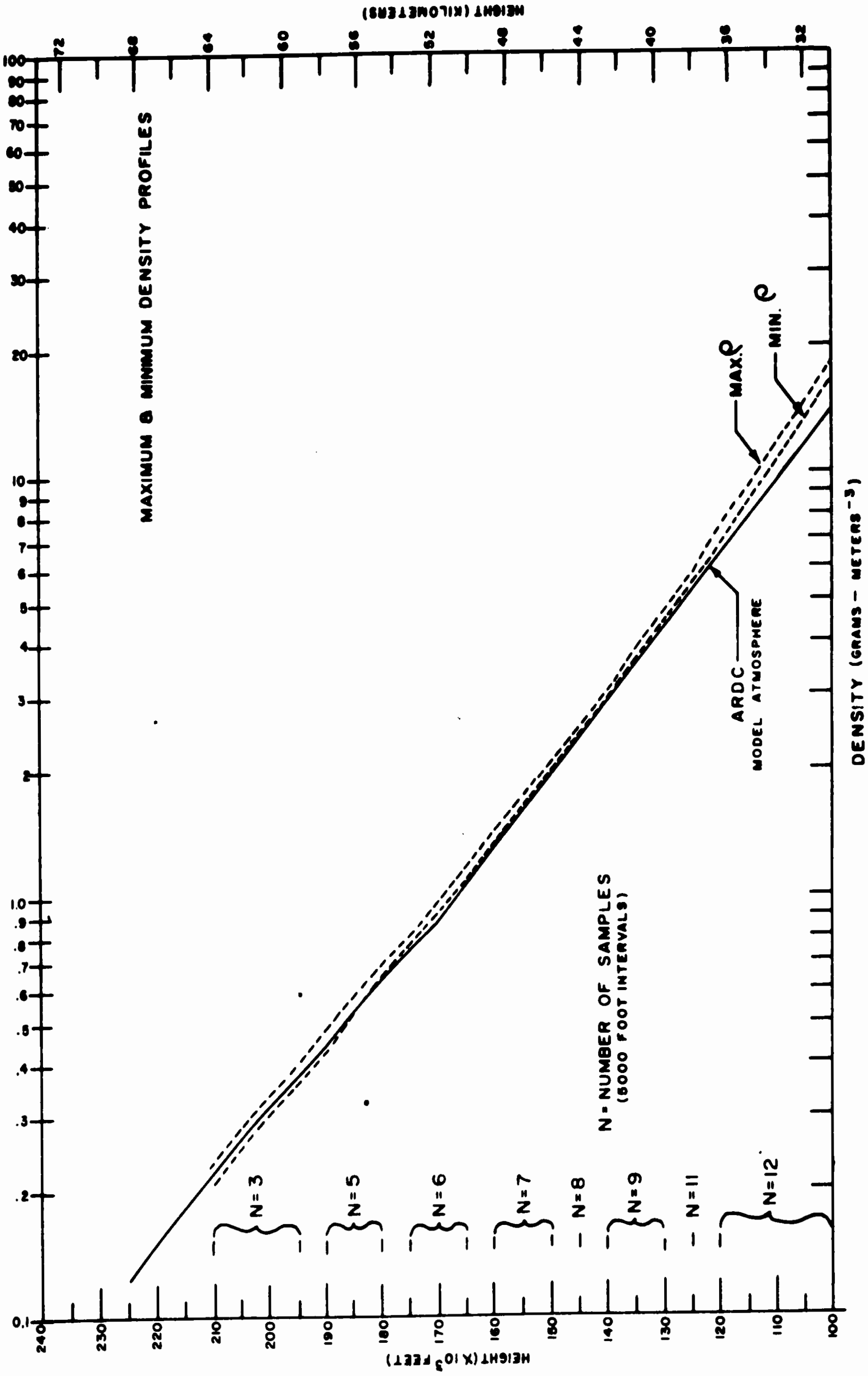


Figure 1 - Maximum and Minimum Density Profiles for Summer 1960, White Sands Missile Range, New Mexico - Latitude 32°

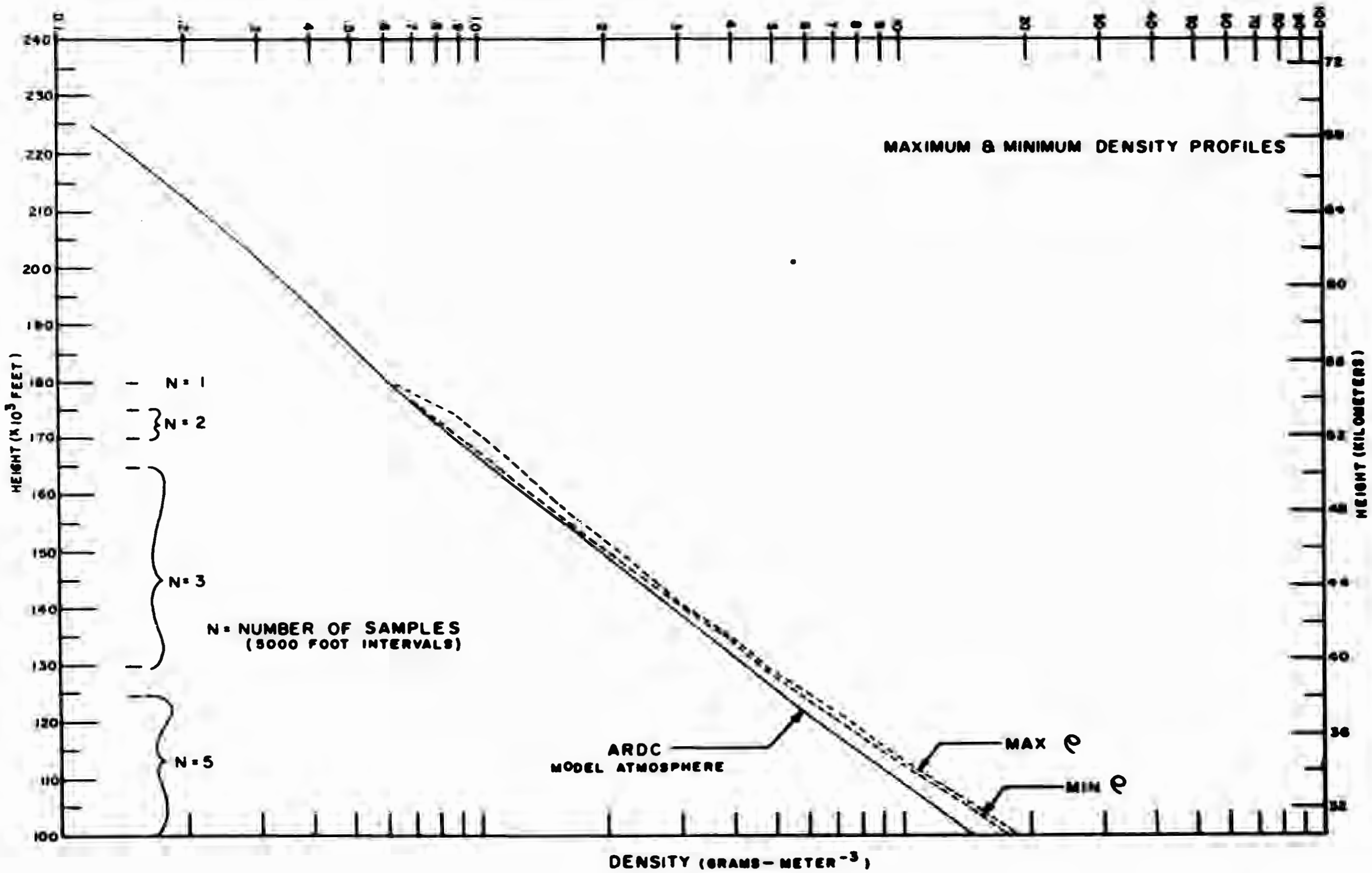


Figure 2 - Maximum and Minimum Density Profiles for Summer 1960, Fort Churchill, Canada - Latitude 58°

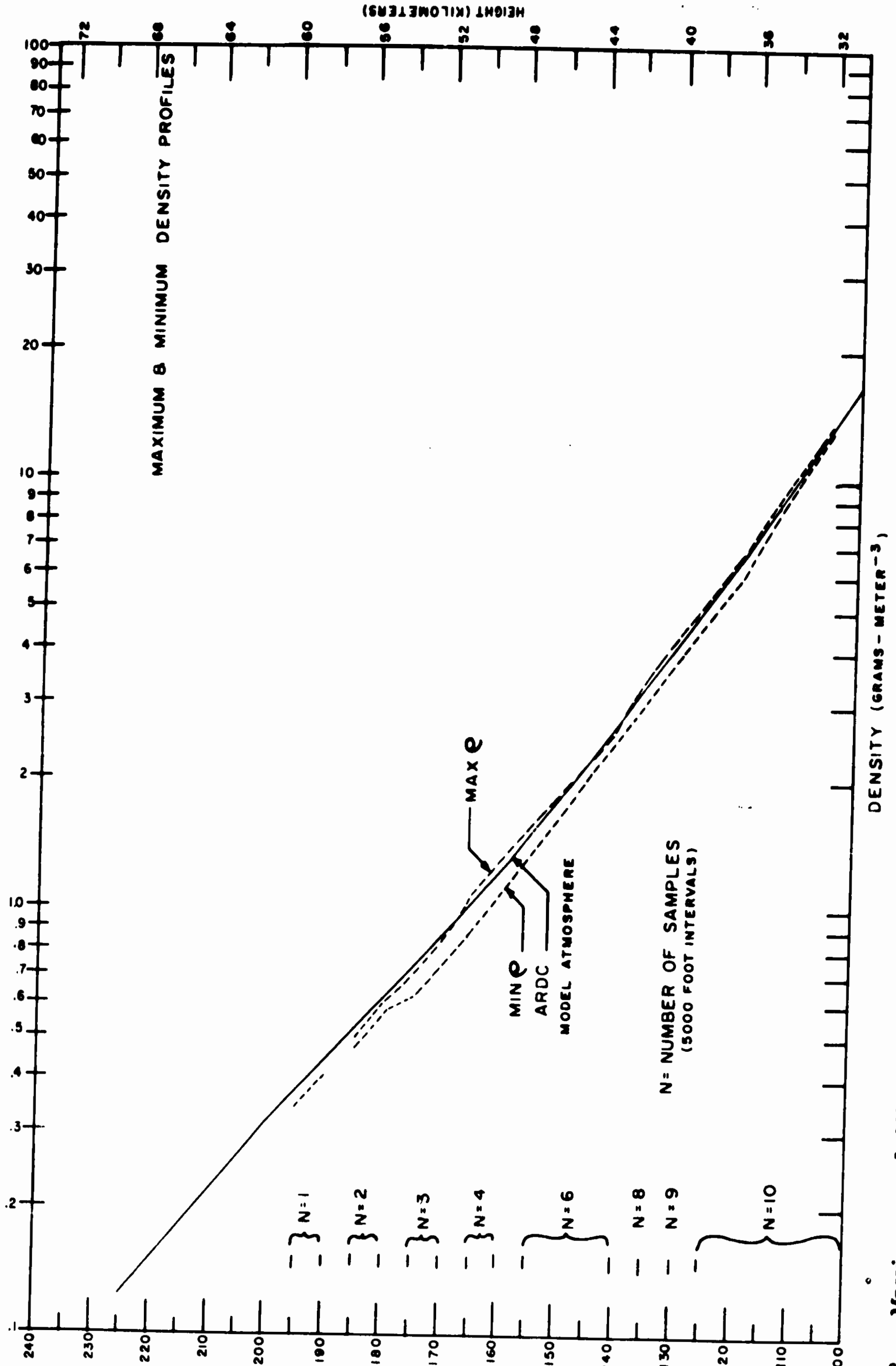


Figure 3 - Maximum and Minimum Density Profiles for Fall 1960, White Sands Missile Range, New Mexico - Latitude 32°

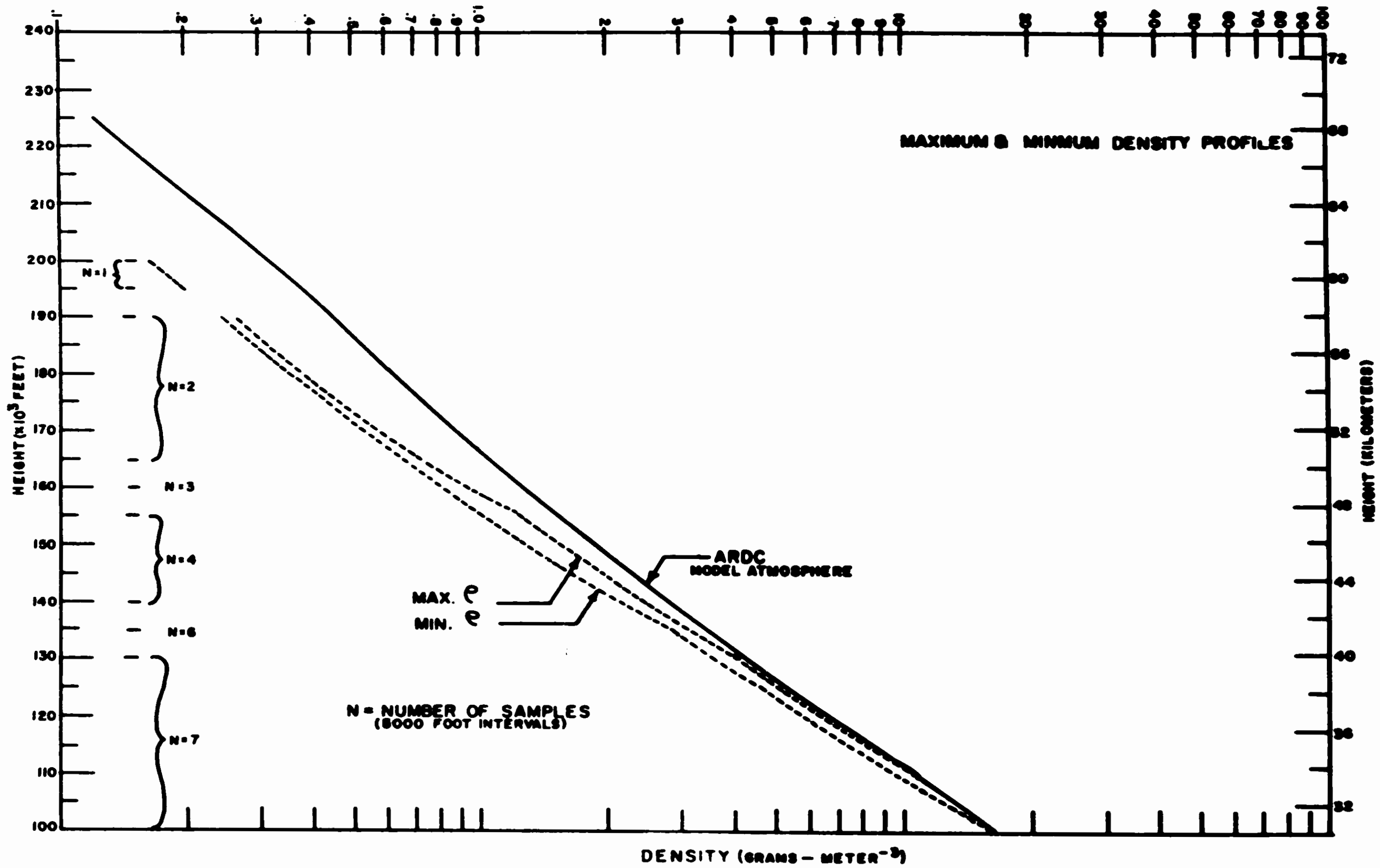


Figure 4 - Maximum and Minimum Density Profiles for Fall 1960, Fort Churchill, Canada - Latitude 58°

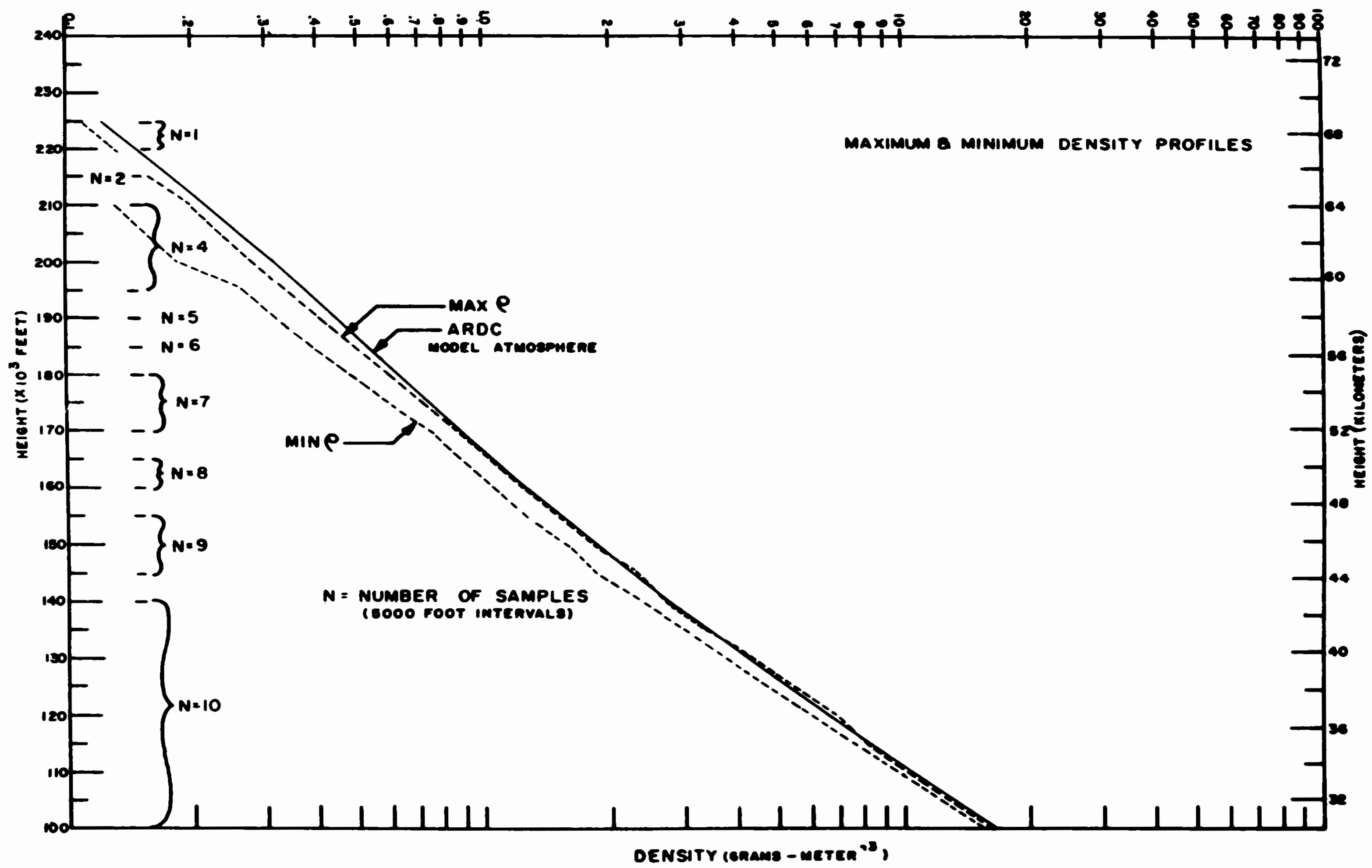


Figure 5 - Maximum and Minimum Density Profiles for Winter 1960, White Sands Missile Range, New Mexico - Latitude 32°

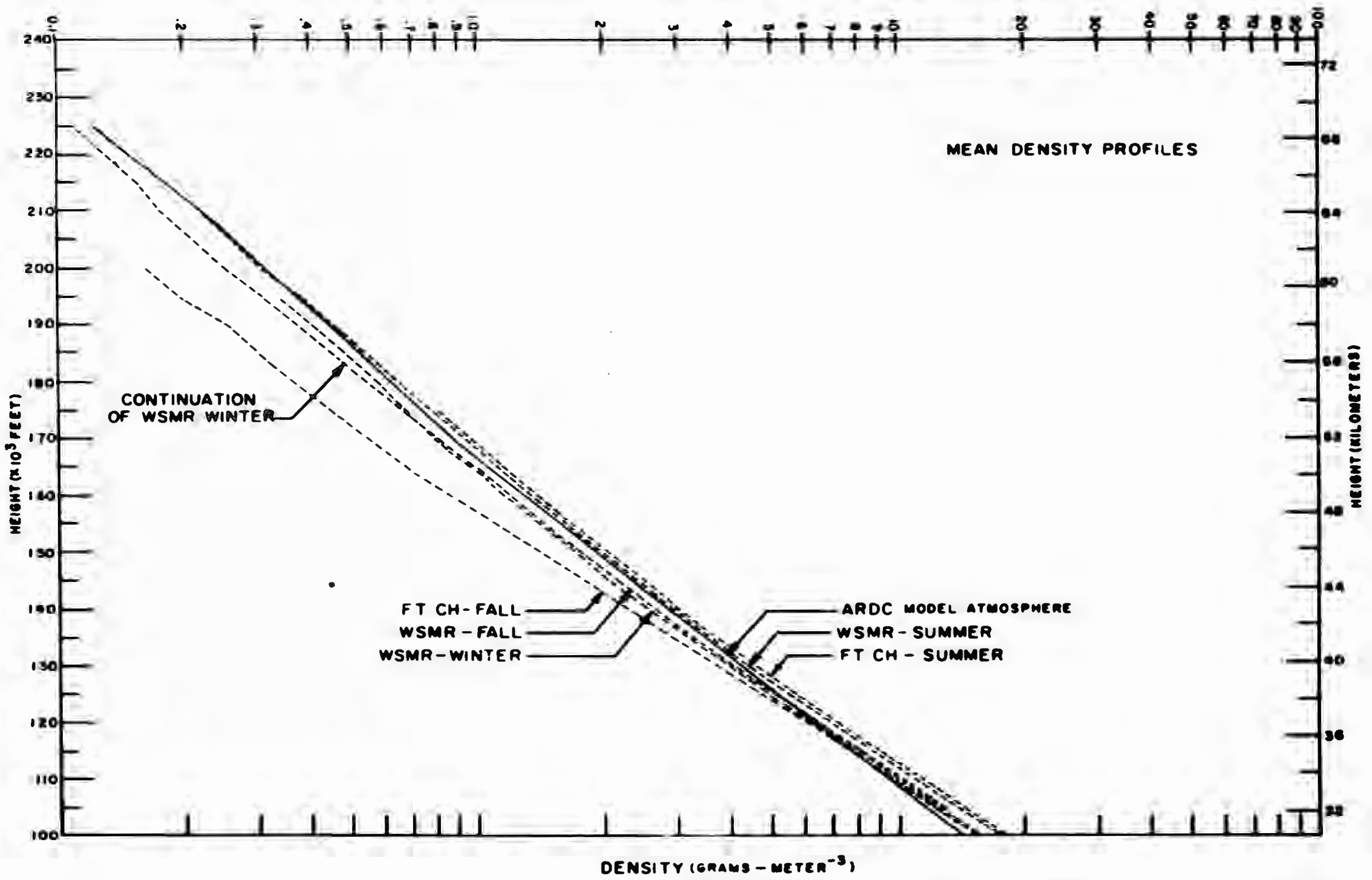


Figure 6 - Mean Seasonal Density Profiles from White Sands Missile Range, New Mexico, and Fort Churchill, Canada

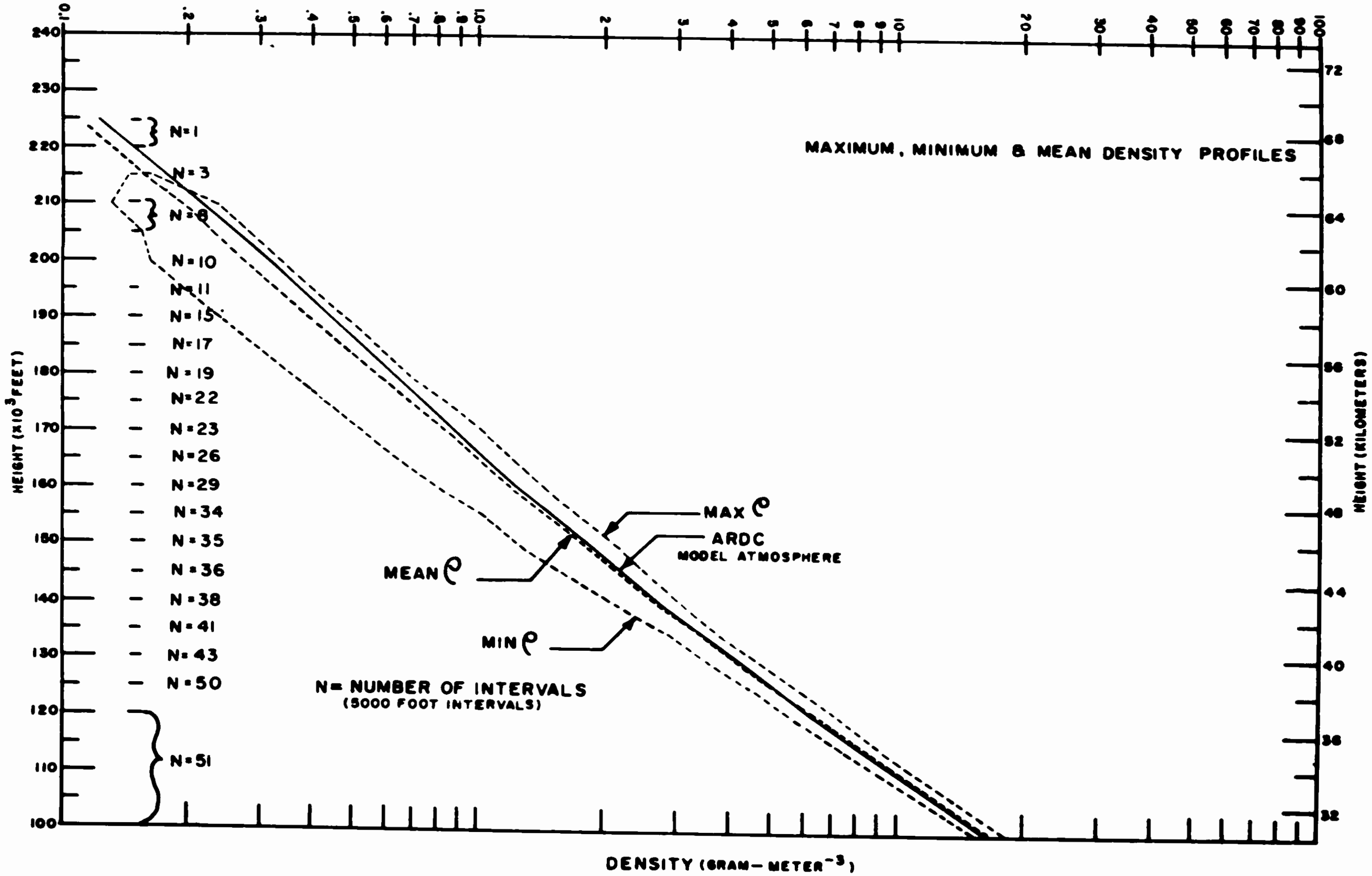


Figure 7 - Maximum, Minimum and Mean Density Profiles of all the Data from White Sands Missile Range, New Mexico, and Fort Churchill, Canada

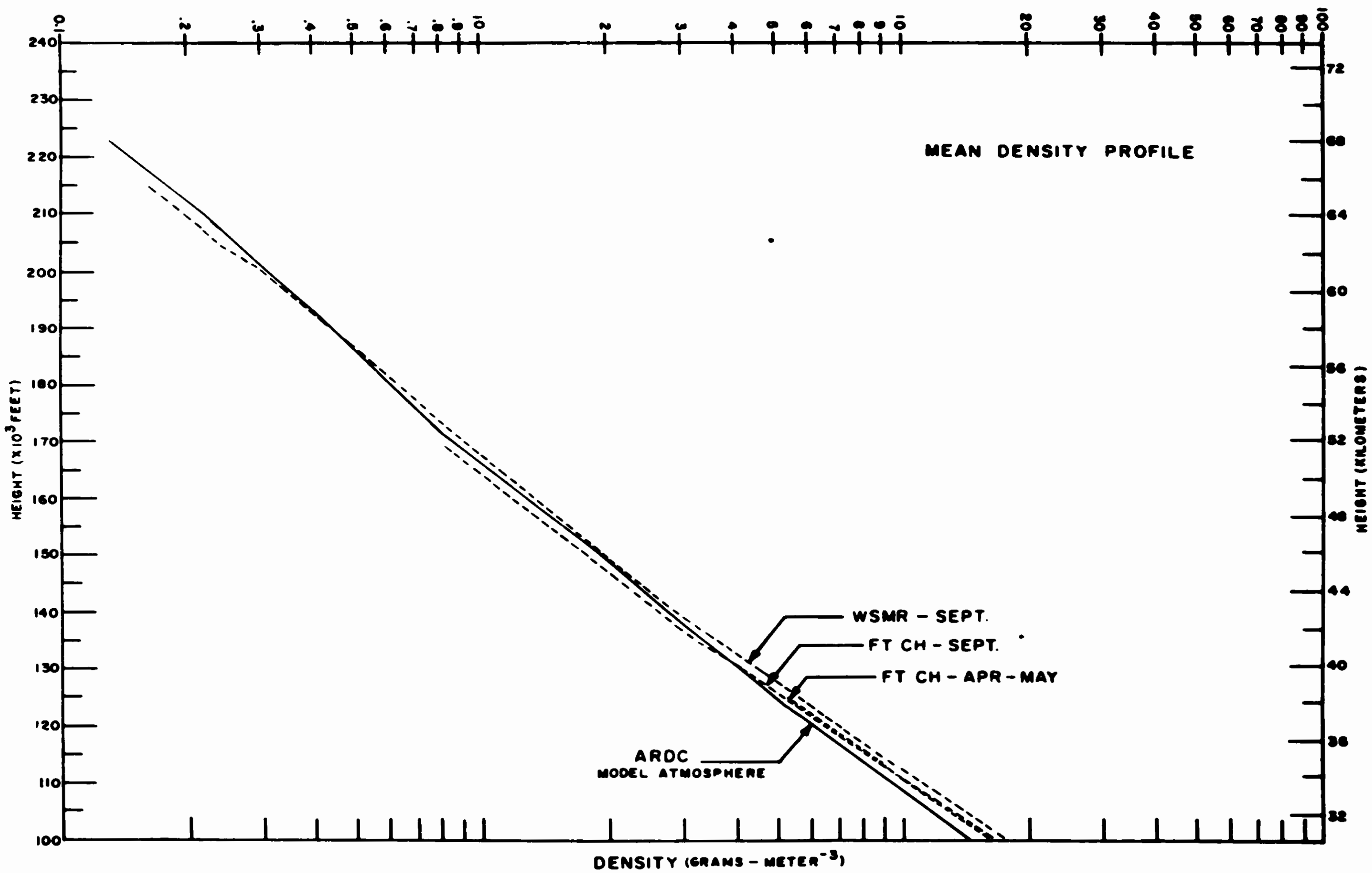


Figure 8 - Mean Density Profiles of Interim Data from White Sands Missile Range, New Mexico, and Fort Churchill, Canada

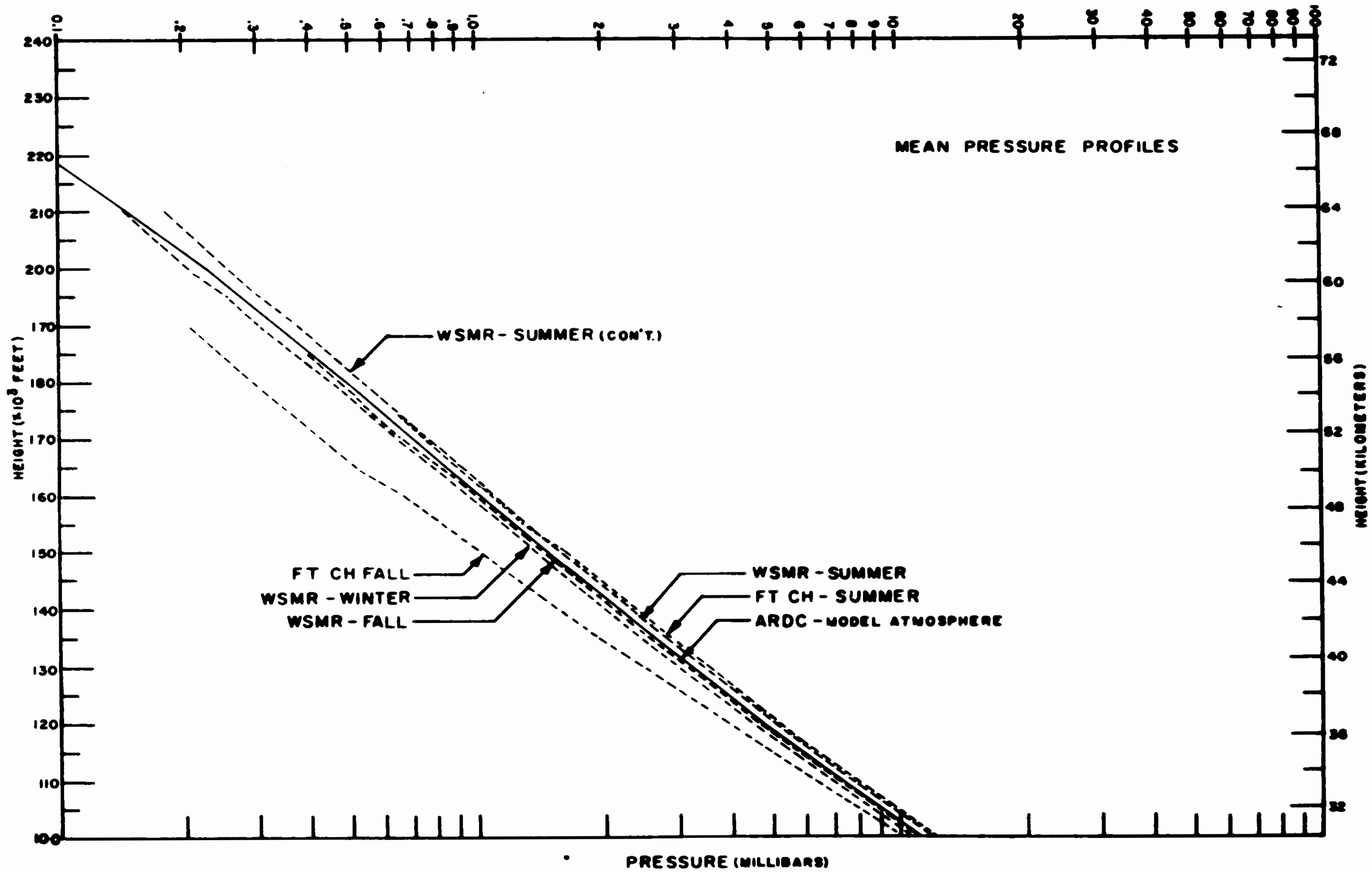


Figure 9 - Mean Seasonal Pressure Profiles from White Sands Missile Range, New Mexico and Fort Churchill, Canada

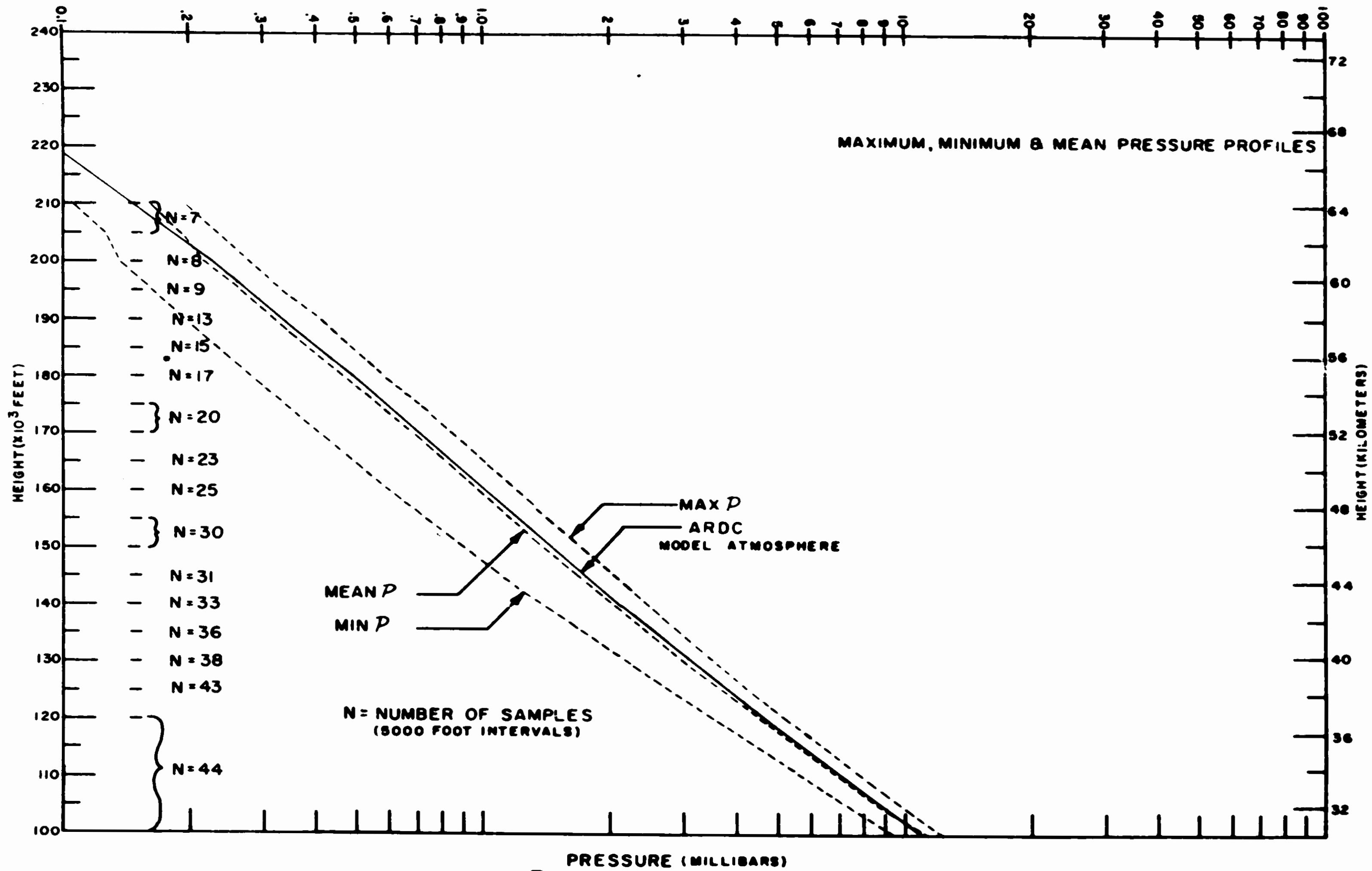


Figure 10 - Maximum, Minimum and Mean Pressure Profiles of all the Data from White Sands Missile Range, New Mexico, and Fort Churchill, Canada

EXPLORING THE ANTARCTIC WITH ICE-SOUNDING RADAR

AMORY H. WAITE,* JR., and STANLEY J. SCHMIDT*
*EXPLORATORY RESEARCH DIVISION "C", INSTITUTE FOR EXPLORATORY RESEARCH, U. S. ARMY SIGNAL RESEARCH AND DEVELOPMENT LABORATORY, FORT MONMOUTH, N. J.

PURPOSE:

This paper presents a brief history of Signal Corps radar ice-depth sounding, discusses past, present and proposed measuring techniques, and details the first ice-depth survey flights ever attempted operationally. These, conducted at thirteen points along 1,000 miles of Antarctic coastline in early 1962 discovered several previously unknown geographical features through 1,000 feet of ice in an aggregate flying time of less than seven hours.

INTRODUCTION AND HISTORY:

Exploratory research, directed towards the determination of the basic factors that govern the behavior of radio waves as they pass through and over the thick ice and frozen terrain of both polar areas, was initiated within the Army by Signal Corps Arctic and Antarctic Research Teams in 1946. Prior to that time much work had been done by civilian scientists across the world, but it was not until large peacetime military expeditions began to organize after World War II that applications of the early findings could be evaluated in the field.

The first large scale research effort, Operation Highjump to the Antarctic, experienced some of the first indications of the excellent dielectric characteristics of thick ice when bare-wire antennas were found to lose little power when dragged behind surface vehicles and others completely buried in the snow proved capable of sending and receiving signals over great distances. Later, the Signal Corps, participating in the International Geophysical Year program, launched eight research projects that produced excellent data relative to high latitude propagation. One of these made operational measurements horizontally through a mile of Antarctic ice for two seasons to clearly demonstrate the radio-wave transparency of the new medium at frequencies from 2 to 400 megacycles,[1]. Another, carried out by men working long periods at low temperatures in hand-

dug, thirty foot deep pits, showed that the value of the dielectric constant of natural ice agreed almost exactly with that obtained from man-made samples in a laboratory 10,000 miles away [1,2]. Others measured ducting just above and below the surface using figures for densities, temperatures, and snowfall obtained on location from glaciological and meteorological specialists. Yet, because the objectives of that period were concerned exclusively with the improvement of electronic equipment intended for cold weather operation, it came about that no Signal Corps polar measurements were made more than a few feet under the surface of the ice until 1957. There was little, if any, interest in phenomena that might exist 5000, 1000 or even 100 feet below.

A change in this viewpoint occurred near the close of the second year of IGY, however, when reports from two widely separate fields of endeavor began to trickle in. Pilots of high latitude aircraft were occesionally reporting serious errors in the height indications of their radar altimeters when operated over thick ice. And, seismologists, traversing some of the earth's millions of square miles of ice-covered terrain, were being considerably retarded by the time-consuming blasting of gun-powder and stretching-out of the long geophone cables indispensable to their sound-wave ice-depth measuring techniques.

Normally, altimeters are a responsibility of the Air Force, and ice-depth measuring systems, though interesting to electronic engineers, are comparatively remote from them in the field of geophysics. In this case, however, the Signal Corps Antarctic Research Team had just proven that ultra-high-frequency radio waves could be propagated several thousand feet through continental shelf ice and was extremely interested in the safety of the Army's small Air Force.

FIRST SOUNDINGS--ALTIMETER ERRORS

A new research program was implemented, therefore, in the austral summer of 1957-1958, whose directed objective was to establish the safety limitations of over-ice radio altimeters for the Army's low-flying cold weather pilots, and, (after verbal discussions with Army Research Office) to investigate the possibilities of a radio ice-depth measuring technique in an attempt to achieve another scientific, though somewhat non-military, goal.

In any event, two slightly modified, 400 megacycle altimeters were obtained and the first attempts to intentionally send radio signals vertically downwards through Antarctic continental ice in January 1958 obtained good depth indications from rocky moraine beneath 500 feet of land-borne ice. These, when corrected with a velocity factor that was based on a widely known value for the dielectric constant of fresh-water ice[2], agreed within five percent of sound-wave measurements made in the same area by trained seismologists. More importantly, indications of approximately 1000 feet of

altitude were registered by the measuring altimeter when mounted on a vehicle less than five feet above the surface. This proved beyond question that similar instruments operating on aircraft, at frequencies of 400 megacycles or lower, could produce potentially fatal accidents and world-wide warnings to that effect were immediately issued. More unfortunately, repeated attempts that season brought no return from the supposedly highly reflective sea-water beneath 800 feet of the Ross Sea shelf near Little America. This, since no other opportunities to measure through floating ice presented themselves until recently, produced one misconception that has caused no little confusion among theorists for several years. After these events the work speeded up and it was soon recognized that the ability to estimate the depth to which a signal of known frequency and power could penetrate, reflect from the bottom, and be received back on the surface again by a receiver of known sensitivity, depended entirely on :

(a) The gradients of dielectric constant and dissipation factor that exist between the ice* surface and the bottom.

(b) The size of the ice crystals that compose the measured mass and might cause signal loss through scattering.

(c) The dielectric constant (usually $\epsilon = 5$ to 10), the wetness and the contour of the sub-ice terrain.

STUDIES IN 1959--FIRST MOBILE SOUNDINGS. DENSITY, DIELECTRIC CONSTANTS, DISSIPATION and VELOCITY FACTORS, and SCATTERING.

The Team moved to the Arctic in 1959 and made several new series of measurements working with the scientists of the Army Engineer Corps' Snow, Ice and Permafrost Research Establishment. These gentlemen, subsequently, furnished invaluable information on subsurface temperatures, pressures, and ice crystal sizes to depths of 1000 feet[3], while their survey teams provided very accurate bottom and surface profiles of a local three mile long ice-ramp that ranged from 100 to 800 feet in thickness[4]. The Team then achieved its second important milestone when it sounded the ramp in less than two hours and found that the radio soundings agreed with the much lower frequency sound-wave depths within the reading accuracy of the instrument employed (i.e., 25 feet) . This established for the first time the speed and accuracy at which ice-depths could be measured with the embryo system. In addition, since their was no apparent deviation between the radio and seismic profiles over the depth

* The term ice is used loosely, but may mean either snow, ne've, or hard ice depending on its specific gravity. This varies from 0.1 to 0.4 for snow, from 0.4 to 0.8 for compacted crystalline ne've and from 0.8 to 0.92 for hard ice which is almost as dense as water. Dielectric constants are approximately 1.6, 2.6, and 3.2 respectively.

range of 100 to 800 feet, the minuteness of the change in dielectric constant from top to bottom was demonstrated. This constant is important since it establishes the velocity at which a radio wave of any frequency will pass through a given medium. For snow, the dielectric constant, epsilon, is about 1.6 and for clear, hard, fresh-water ice is approximately 3.2. Similarly, the ramp survey showed that the characteristics of the underlying terrain were such that they reflected both low-frequency sound-waves and ultra-high-frequency radio waves equally insofar as the velocity factor was concerned. This, either in ice or rock, varies inversely as the square root of the dielectric constant just discussed. The existence or non-existence of water at the ice/rock interface is unknown, although the low temperature present, -10°C, denies this possibility. In this connection, it must be stated that several glaciologists 5 dispute the existence of melt-water beneath glaciers, and that temperatures in polar ice masses support this theory since they are several degrees colder than those necessary to cause melting at existent depths and pressures.

A series of substitutions of known values of antenna circuit attenuators also provided data on dissipation factor characteristics which allowed the authors to prophesy that the 400 megacycle test system should reach a specific depth of 1230 feet. This was later borne out in practice in three separate ice masses within 2 percent of the figure calculated. The dissipation or loss factor is a measure of the goodness of any dielectric which controls the attenuation suffered by waves passing through it; for ice the figure of 0.001 is representative. This factor varies almost directly with temperature and density and inversely with frequency. No ice has been observed in either the Arctic or Antarctic in which the crystal size was large enough to cause loss from scattering. Such losses can therefore be disregarded until much higher sounding frequencies are employed.

BASIC SYSTEM AND ALTIMETER CORRECTION FACTORS

Figure 1 clearly outlines the basic elements of the radar sounding system. Pulses, at the rate of 100,000 per second, sent downwards through the ice to bounce off the terrain below, return to the receiving antenna and thence to the receiver. Some energy is reflected from the surface so that should the system be airborne, subtraction of the surface indication shown on the oscilloscope (Figure 2) from the bottom indication will give the thickness of the mass being surveyed when corrected for the in-ice wave velocity (0.53-0.59). This cannot be accomplished from a surface vehicle because the ground pip on the scope spans 250 feet. A typical 1000 foot deep sounding situation is indicated on the scope as 1725 feet and the losses suffered by the radio wave are approximately 70.0 db because of "free space loss" down and back, 40 db from "absorption" in the ice itself, about 0.8 db "refraction losses" entering and leaving the air/ice/air interface, plus approximately 15 db at the

ice/rock junction. This totals 125.8 db in a system that has an overall gain of 130. The resultant signal level is 10 microvolts which is ample for a good scope pip indication.

The project thus reached a state by the spring of 1960 where curves could be drawn that anticipated depth-measuring signal losses in ice at frequencies from 50 to 4000 megacycles and at depths up to 30,000 feet[6]. Quickly applicable correction factors, with which the height indications normally registered by ice-sounding altimeters could be converted to read ice-depth, were also available. For example, the wave velocity in any medium is, C, the velocity in free space, divided by the index of refraction of the medium, n. Since n is equal to the square root of the dielectric constant, epsilon, by definition, and epsilon for hard ice is approximately 3.2, as stated above, we find the wave velocity in ice to be $\frac{186,000}{\sqrt{3.2}}$ or 104,000 miles per second. The altimeter correction factor is thus $\frac{104,000 \text{ (ice)}}{186,000 \text{ (air)}} = 0.56$ or 56%. Scope indications, therefore, multiplied by this figure give ice depths directly in feet.

FIRST SOUNDINGS IN FLIGHT

The Team returned to the Arctic in the summer of 1960 with lower frequency and more powerful units that were expected to penetrate much thicker ice than the earlier 400 megacycle ones. It then made several new series of measurements which, briefly summarized, showed that (a) 100 megacycle systems with commensurate antennas should be adopted for future sounding activities; (b) that altimeters operating at 4300 megacycles were much safer than 400 megacycle ones, but even the higher frequency units penetrated the surface as much as forty feet in summer and probably farther in winter when it was not wet with melt-water; and (c) that the modified 7 watt, 400 megacycle, pulsed Radar Altimeter SCR-718, still used in many present-day aircraft and throughout all early ice-depth studies, could be relied upon for soundings in solid blue ice up to 1200 feet thick. Radio ice-depth sounding equipments were also flown in helicopters that season to prove that the bottom could be outlined beneath ice as much as 1000 feet thick at high speeds.

ANTARCTIC PLAN-1962

Only a few successful radio soundings through sea-borne ice and a series of comparisons with seismic data previously acquired by the National Science Foundation in Antarctic ne've, were needed by July 1961, to complete the project's list of required basic data. The Team, therefore, joined the Navy's Operation Deepfreeze and sailed to the Antarctic on the U. S. S. Glacier in November 1961. It planned to:

(a) Test a high power, low frequency unit that might reach the bottom of 4000 or 5000 foot deep ice where National Science

Foundation (NSF) data had already been taken so that comparisons would be possible;

(b) Measure through floating ice and thus prove or disprove the 1957 reflection failures;and

(c) Compare readings made through thick ne've, with those of NSF, to achieve the dielectric value that was needed to complete the data curves.

Then, if the deep-sounding 100 megacycle equipment operated satisfactorily, long flights into the interior would be arranged. If not, the Team would accompany the U. S. S. Glacier on its proposed 1000 mile exploration cruise along the Marie Byrd Land coast to evaluate a 400 megacycle system under operational conditions for the first time.

EQUIPMENT

The equipment needed for the expedition was assembled and throughly tested before departure from the United States. It included:

(a) Two 400 megacycle altimeters, with 100 percent spares, which were taken for all measurements on or near the Antarctic coastline. They were mounted on single, two foot square, wooden panels to facilitate quick installation in either aircraft or surface vehicles. Their 4 foot square anternas were split and shipped in special cases to permit rapid assembly on location;

(b) Two 4300 megacycle units which were mounted similarly, but with their antennas arranged for quick placement outside of an open helicopter door; and

(c) A newly designed 100 watt 100 megacycle system equipped with high gain Yagi antennas, and a highly sensitive receiver.

Preliminary calculations showed that this system should receive reflections back from the bottom under 5000 feet of ice, the 4300 megacycle systems should prove adequate as surface indicators to permit rapid and accurate profile work, and the 400 megacycle units should be able to sound any landborne ice up to 1200 feet in thickness without difficulty. Later, it was found that the Glacier's helicopters were already equipped with 4300 megacycle units so the Signal Corps sets were used only as spares. The 4300 megacycle system was found to have an average penetration of 25 feet into most of the Antarctic ne've flown over in 600 miles of testing, but there were occasions when badly deceptive altitudes were observed. These have caused the authors to advise against reliance on equipment of this frequency over thick ice until further study has been undertaken. The 100 megacycle system was tested throughly on

three separate occasions in areas where a very weak signal should have brought results, but no signal was received. The equipment has, therefore, been set aside for future improvement though the principle and deep sounding objectives remain unchanged. The two 400 megacycle units continued operating without fail throughout the trip, as did the laboratory constructed 12 volt DC to 115 volt 400 cycle AC power supplies. One system has now been loaned to the Navy for the training of future survey teams. Such training for a smart high-school graduate should make him a radio surveyor in 10 days, but the technical maintenance of the equipment requires trained repairmen.

ANTARCTIC SURVEY FLIGHTS-1962

The authors, with the outstanding cooperation of Captain Edward Grant, USN, the officers and men of the U. S. S. Glacier, and Lcdr. Ernest Burks, USN, the pilots and their crews of Helicopter Squadron-4, NAS, Lakehurst, N. J., completed a total of thirteen radio sounding flights between December 1961 and February 1962. These and the results they made possible form the major interest of this paper. They started in McMurdo Sound and ended 1000 miles to the Eastward as shown by the map in Figure 3.

Flights 1 and 2 were made south of Ross Island on thick floating ice (Fig. 4) and immediately established much stronger reflections from sea water than had ever been received from underlying rock. On Flight 2 a line of flags four and a half miles long, placed by New Zealand scientists to measure ice drift, was sounded at the rate of 5 minutes per flag. Depths varied from 560 to 620 feet and were later verified by a second run. This area was 20 miles from the ship and required a round trip of 45 miles.

Flight 3 traveled 50 miles from the ship to prove Butter Point to be 400 feet deep at the famous British gasoline cache, and then surveyed the three mile wide Ferrar Glacier with 5 soundings to show its average depth to be approximately 190 feet. This is much less than was previously believed possible and indicates that the glacier is probably decadent. The soundings of this flight and also 1 and 2 were made by lifting the plate-type antennas out of the helicopter (Fig. 5) and placing them directly on the surface. Results were clear and instantaneous. No thick ice was found eastward of Ferrar where the glacier edge floated,as shown on current Hydrographic Office Chart #6666.

Flights 4, 5, and 6 were made (Fig. 6) with the plate antennas mounted fore-and-aft under the body of the helicopter for the first time so that landings were no longer necessary. This procedure greatly speeded up the sounding process, which was simplified further when Flights 4 and 5 proved that laborious barometric altimeter readings could be elininated when pips were registered on the scope that adequately profiled the surface. Surface contours

could not be achieved without some means of establishing the plane's absolute elevation above sea-level, but accurate barometric altimeters carried to the Antarctic for the purpose were found unsatisfactory for mapping. They do provide reasonable estimates if carefully calibrated at frequent intervals, however.

Flight 4 (Fig. 7) tested the 100 Mc equipment at NZ Flag #4 south of Ross Island and proved it inoperative. Attenuator substitution tests on the 400 Mc unit then undertaken at this point gave a depth of 580 feet and a reserve signal of 29 db. This again proved the unit capable of a total depth of 1230 feet.

Flight 5 (Fig. 7) surveyed the northeastern tip of White Island with four parallel 5 mile legs which definitely established a hitherto unknown ridge. Depths varied from 200 to 450 feet and agreed very closely with seismic soundings made earlier by Crary 7.

Flight 6 (Fig. 7) surveyed 80 miles of the 10 to 40 foot thick McMurdo Sound bay-ice in an attempt to find a thick enough area for a new base on which to land wheeled aircraft. This was made at the request of Rear-Admiral Schinn of the Navy. No suitable landing area was found, but the experience gained on this long, low flight made it worth while and acquainted both the Navy and the United States Antarctic Research Project personnel with the Signal Corps' new capability. Several conferences with scientific personnel of both the American and New Zealand groups outlined several tentative sounding areas which might prove profitable to the Team. One of these, the Tasman Glacier near Mt. Cook in New Zealand was not sounded as planned during the Christmas Holidays because no air transport could be obtained by Victoria University and New Zealand's Department of Scientific and Industrial Research. Nevertheless, as the ship moved eastward in January 1962, several U. S. Antarctic Research Project suggestions were followed. Flights 5 and 6 belong in this category as did-

Flight 7 (Fig. 7) which sounded the Ross Shelf ice east of Cape Crozier on Ross Island. A strip five miles long was flown in 7 minutes with excellent results that showed the ice comparatively thin near the warm volcanic island and tapering downwards from 220 to 330 feet thick in five miles. Accurate measurements from the bridge of the ship showed the ice above water to be slightly more than 16 percent of the total mass as experienced by earlier geophysicists. These height values added weight to the radar soundings obtained, which were taken every minute as under normal conditions. As the helicopters invariably flew at a speed of 60-65 knots, this worked out as one reading per nautical mile.

The Ross Shelf ice thickened rapidly eastward of the area where Flight 7 was made (Fig. 8) and continues along most of its 350 mile width at a thickness that varies between 800 and 1000 feet.

Flight 8, intended to rechart the area near Admiral Byrd's Original Camps on the shores of the historic Bay of Whales, surveyed 54 miles in 57 minutes (Fig. 9). Depths not only agreed very closely with those reported by Poulter in 1953[8], but the inland end of this flight definately indicated the demarcation line between sea-borne and land-borne ice at the northern edge of Roosevelt Island and also clearly registered every large ice canyon that the helicopter passed over. The ice was definitely thinner where it rose up the northern slopes of the island. This flight was beyond question the most interesting one of the expedition since it not only proved the early findings which had been questioned by many recent authorities, but also rediscovered Byrd's first Little America. Towers constructed 34 years ago were still showing above the drifts.

Flight 9 will be discussed with Flight 13 later.

The ship left the Bay of Whales, visited the IGY Little America at Kainan Bay, briefly, then proceeded to Okuma Bay, Cape Colbeck on Edward VII Peninsula, and then sailed eastward to explore almost 500 miles of the Marie Byrd Land Coast where no ship had ever been before. The three flights made after the visit to Cape Colbeck were carried out at the specific request of the Task Force Commander who was accurately mapping the area.

Flight 10, (Fig. 10) intended to locate the dubiously mapped edge of the continental shelf ice off the Rupport Coast, flew 19 miles from the ship inland in poor visibility, turned eastward for 4 miles, and then swung back to the ship again, having mapped over forty miles of the surface and the bottom in forty minutes. The edge of the 1000 foot thick shelf was clearly defined and so were the edges of three massive icebergs. Clear delineation between the shelf and a large submarine rock ridge which pushed up into it for several hundred feet was also observed. This flight recorded observations at such short intervals (15 seconds) that they outlined the effects of melt on the bottom of the seaward edge of the shelf for the first time. This degree of melt has been of interest to glaciologists for many years. There was also evidence that one of the three bergs seen was grounded showing the presence of another rocky pinnacle. Lastly, it was very interesting to note that the shelf seemed to thicken inland of the anchoring ridge which may with study prove to be an indication of the shelf's age. Obviously, the forward motion of the 1000 foot thick shelf (weighing 990,000,000 tons per square mile) was being slowed up by the rock ridge. This flight also sighted a mile-long rocky island standing 600 feet above sea level and ten miles SE of the southwestern turn.

Flight #11 (Fig.11) zigzagged on five mile legs across the mapped edge of the Getz Ice Shelf to attempt to draw the actual coastline beneath 1000 to 1300 feet of ice. It was very successful

and will be used to change existent maps. Ten miles of the coast was mapped in less than 10 minutes while a total of 65 miles was flown in an hour to get this data.

Flight #12 (Fig.11) was the last flight made on the eastern leg of the journey because the ship was ordered home for emergency ice-breaking in McMurdo Sound when Mt. Siple, a major objective, had just become visible on the horizon 97 miles away. A snow mound, 15 miles from the ship, appeared at this time to be an island, but was doubted by the navigators. Permission to make one last flight was asked by the discoverer and granted by the Glacier's Captain, and a four mile wide island was both proven and profiled in 15 minutes flying time. Its contour is clearly shown in Figure 11 which agrees exactly with soundings accomplished by the ship's sonar nearby.

The ship then returned rapidly to Kainan Bay where Flight 13 reran most of the surface observations made earlier in Flight 9, when East Cape, Camp Coldbottom and Little America were sounded accurately at points where Crary[7] had made his seismic survey in 1957. Another short traverse was made out over the Byrd Land trail area to compare in-flight to on-surface radar depth soundings. This flight produced figures that agreed within 10 feet of Crary's measurements in several places and exactly matched those acquired by actual drilling by SIPRE3. They also allowed a definite determination of the dielectric constant and loss factors to be used in 900 foot thick ne've' and these have been employed for all computations in the last chapter of this paper.

CONCLUSION

The capability of sounding through ice, with radio waves, in either the Arctic or the Antarctic, is now a proven fact. Slightly modified standard stock items, already installed on many aircraft, can be utilized to map sub-surface features through up to 1400 feet of ice by alert enlisted men with brief training as fast as aircraft can fly. The project has reached a point where immediate action can be taken to perfect the means and thereby to increase attainable depths. The basic research is practically complete.

The flights reported herein surveyed more than 300 linear miles through ice that ranged from 200 to 1300 feet in thickness in less than 7 hours total flying time and only two men were involved.

The authors believe the immediate future will witness 5000 foot deep soundings being taken by over-ice pilots as part of normal flight routine while recorders roll the data into compact cartridges for later scientific review. Greater depths are similarly within view, but these may require as much as two or three seconds delay for each sounding.

BIBLIOGRAPHY

1. Waite, A.H., "Ice Depth Soundings with Ultra-High Frequency Radio Waves in the Arctic and Antarctic, and Some Observed Over-Ice Altimeter Errors", U.S. Army Signal Research and Development Laboratory Technical Report 2092, 16 December 1959.

2. Von Hippel, A., editor, "Dielectric Materials and Applications", the Technology Press of MIT and John Wiley and Sons, Inc., New York, page 301, 1954.

3. Ragle, R.H., Hansen, B.L., Patenaude, R.W., and Gow, A.J., "Deep Core Drilling in the Ross Ice Shelf, Antarctica, U.S. Army Snow Ice Permafrost Research Establishment Technical Report 70, June 1960"

4. Roethlisberger, H., "Private Correspondence Concerning Densities of Ice and Temperatures at Different Depths in Arctic and Antarctic Regions, 22 December 1960".

5. IGY Bulletin of the National Academy of Science, "Glacial History of McMurdo Sound Region, Antarctica", page 1, No. 36, June 1960.

6. Waite, A.H. and Schmidt, S.J., "Gross Errors in Height Indication from Pulsed Radar Altimeters Operating Over Thick Ice or Snow", 1961 IRE International Convention Record, Part 5, pages 38 - 54.

7. Crary, A.P., "Glaciological Studies at Little America Station, Antarctica, 1957 and 1958," Chap. 3, pages 25 - 61.

8. Poulter, T.C., "Seismic Measurement on the Ross Shelf Ice", Trans. of the American Geophysical Union, 1947, Vol. 28, Nos. 2 and 3.

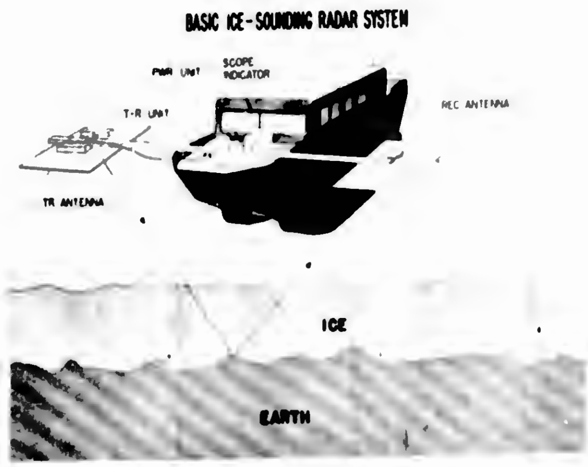


Fig 1. Basic Ice-Sounding Radar System

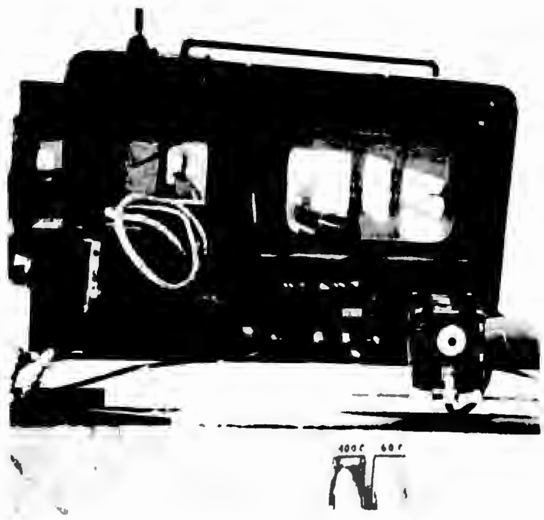


Fig 2. Equipment Installation on Weasel,M-29-C

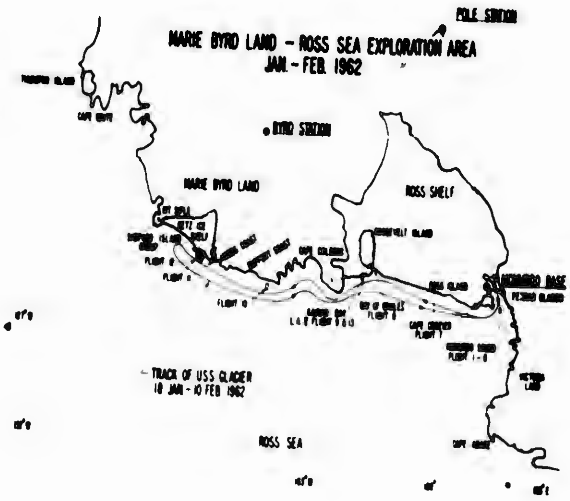


Fig 3. Chart of Exploration Area, 1962

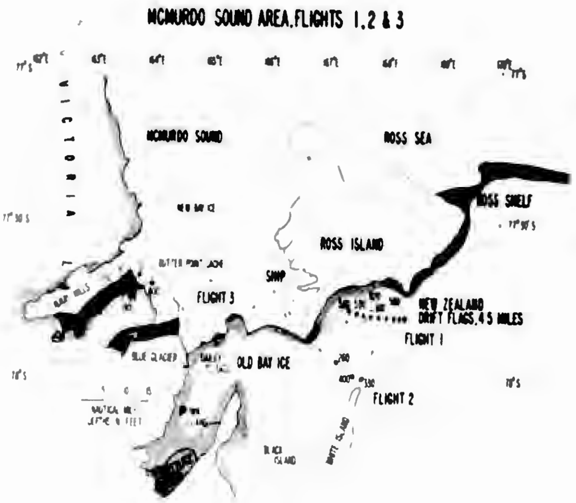


Fig 4. Chart of Flights 1, 2 and 3

Fig 5. Sounding on the Ferrar Glacier

Fig 6. Antenna Installation on Helicopter, HO4S

Fig 7. Chart of Flights 4, 5, 6 and 7

Fig 8. Typical Ross Shelf Ice

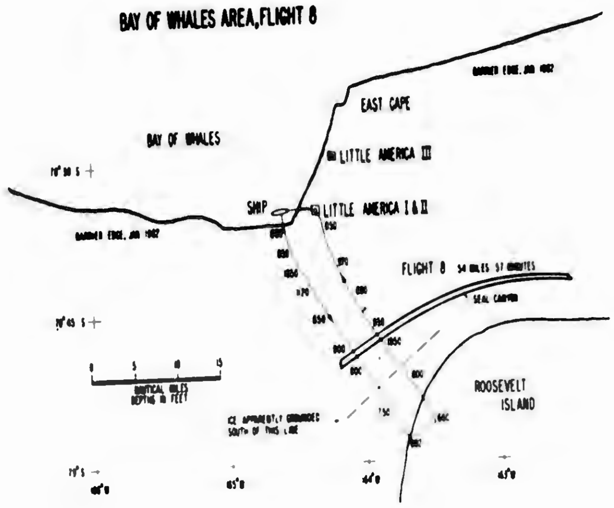


Fig 9. Chart of Flight 8

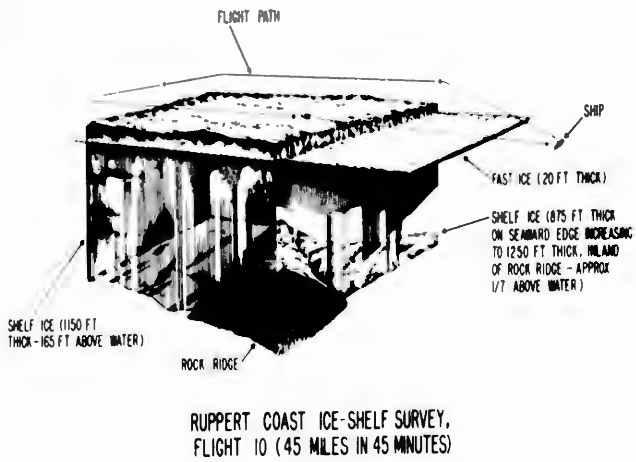


Fig 10. Chart of Flight 10

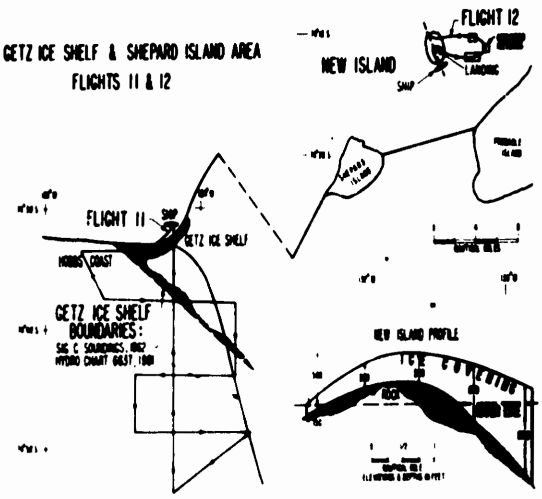


Fig 11. Chart of Flights 11 and 12

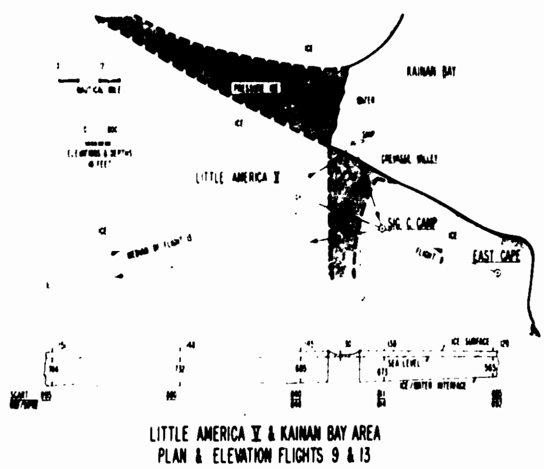


Fig 12. Chart of Flights 9 and 13

A MATHEMATICAL MODEL FOR A BALLISTIC ROCKET

DR. EVERETT L. WALTER
U. S. ARMY SIGNAL MISSILE SUPPORT AGENCY
WHITE SANDS MISSILE RANGE, NEW MEXICO

I. INTRODUCTION

In designing, testing, or firing a rocket, the use of some type of mathematical model which approximates the rocket's flight will result in a savings in time and cost. We present such a model for a ballistic rocket (i.e., a rocket for which there is no guidance after it is launched). This model was recently developed and is presently used at the U. S. Army Signal Missile Support Agency (USASMSA) at White Sands Missile Range (WSMR) in support of firing operations. It consists of a set of simultaneous differential equations which can be numerically solved rather quickly on a high-speed computer, and has as its output, information which accurately describes the hypothetical trajectory of the rocket. By a hypothetical trajectory we mean the trajectory which would result if the rocket were perfectly aligned and if the rocket and atmospheric data were exact. Because of the generality of the equations, there is no hope for an analytical solution; hence, a high-speed computer is necessary to integrate them numerically with sufficient speed and accuracy.

In the design of a mathematical model it is necessary that the hypothetical trajectory have accuracy sufficient for the intended purpose, but at the same time the model must not be too complex for the available computing facilities, nor so complex that the time and cost are excessive. One would desire also that the model be flexible enough to permit its application to many different types of rockets in any foreseeable environment. In many instances, it is desirable to have different models, each intended for a specific purpose.

It is felt that the model presented here is adequate for the support of the firing of most ballistic rockets provided that neither the range nor the height exceed 300 miles. Its original purpose was to predict accurately the impact point of a rocket when its flight parameters and the atmospheric data were given. That the model had to

be somewhat detailed, is seen by the fact that the effect of the winds alone upon the impact point of an Aerobee-Hi rocket has been known to be greater than 50 miles. Since well over 100 Aerobee-Hi rockets, as well as countless other types of rockets, have been fired at WSMR alone, the importance of a good, working model for ballistic rockets is apparent.

Such models as indicated in (1) or (2) have been used successfully at USASMSA in the past, and still serve their purpose. There are many models more accurate than these, but each has the disadvantage of being more complex. We present such a model in this paper.

In Section II, the equations of motion which describe the flight path of the rocket are given without proof. How so-called "perturbation equations" can be derived easily for various purposes and how they can be used to accelerate certain calculations are indicated in Section III. The numerical solution of the equations of motion and the perturbation equations on a computer is explained in Section IV. In Section V we discuss how the model could be used in the support of an actual firing. Finally, Section VI indicates the future applications planned by USASMSA. Computation time required for the model as well as the amount of personnel time required to write the program are indicated also.

II. THE EQUATIONS OF MOTION

Many attempts have been made to set up equations of motion with the thought of solving them analytically. All of these must make some assumptions on the properties of the rocket and the atmosphere which will cause some errors in the resulting hypothetical trajectory. In many applications, especially for the larger rockets, these errors are intolerable. To minimize these errors, we shall present a set of six equations of motion in which very few assumptions or approximations have been made. This, of course, implies that these equations (which are simultaneous, second-order differential equations) must be solved by numerical integration (and certainly on a high-speed computer in order to get results within a reasonable time and yet retain the desired accuracy). These six equations describe the position in three dimensions, the attitude, and the spin rate of the rocket. Because we do not insist upon an analytical solution, we may include the effects of the jet force due to the main rocket motor (considering variable atmospheric pressure); the jet damping due to this force; the moments due to spin rockets; the aerodynamic forces (i.e., drag and lift); the moment due to the aerodynamic forces (called the restoring moment); the moment created by canting of the fins; gravity, using the inverse square law, and considered not as a parallel force, but as a central force; Coriolis force and the centripetal force due to the rotation of the earth; the apparent thrust and moment due to loss of mass; the force and moment due to a changing angle of attack (the so-called "cross-spin" force and moment); the winds; the gyroscopic effects of spin and of tumbling; a variable speed of sound; a variable air pressure; and other effects.

No proof or development of the equations of motion will be given here. For a detailed development, see (3). Most of the necessary physics can be found in such texts as (4). Before writing the equations we shall give all of the necessary definitions and notation.

(x,y,z) - The position coordinates of the center of gravity of the rocket relative to the launcher; x east, y north, z vertical (as measured by a plumb bob at the launcher).

θ - The zenith angle (in radians) of the rocket axis.

α - The azimuth angle of the rocket axis measured counterclockwise from the x axis.

$\dot{\beta}$ - The rate of spin of the rocket about its axis; positive if the spin vector is directed (using the usual right hand rule) along the positive rocket axis (i.e., from tail to nose).

C_p - The distance from the nose to the center of the aerodynamic pressure (a function of Mach number).

C_g - The distance from the nose to the center of gravity (a function of time).

(u_x,u_y,u_z) - The velocity vector coordinates of the wind, indicating (to be consistent with standard meteorological practice) the direction from which the wind is blowing.

g_o - The measured value of gravity under some average point in the trajectory (such as under the expected peak) which has been corrected to sea level.

λ - The latitude of the launcher.

R_o - The sea-level radius of the earth at the launcher.

ω - The angular velocity of the earth about its axis.

ρ - The density of the air at the present rocket position.

d - The diameter of the rocket.

h - The length of the rocket.

C_D - The coefficient of drag (defined by the equation $D = C_D \rho v_a^2 d^2$ where D is the force due to drag and v_a is the velocity of the center of pressure relative to the surrounding air). The value of C_D is given as a function of Mach number.

C_L - The coefficient of lift, defined by the equation $L = C_D \rho v_a^2 d^2 \sin\delta$, where L is the lift force (i.e., normal to the velocity vector, v_a) and δ is the so-called angle of attack, defined later in the paper.

S - The spin moment created by canting of the fins which is a function of the size and shape of the fins, the amount and

direction of the cant, Mach number and ρ. The sign of $\dot{S}$ is positive if the canting tends to increase $\dot{\beta}$ algebraically, and otherwise is negative.

$T_{s.l.}$ - The thrust of the main rocket motor plus the forward thrust of any spin rockets, measured at an atmospheric pressure of $P_{s.l.}$ (possibly a function of time).

P_a - The atmospheric pressure (a function of rocket position).

$S_{s.l.}$ - The spin moment due to spin rockets measured in an atmospheric pressure of $P_{s.l.}$ (possibly a function of time).

s - The sign of $S_{s.l.}$, positive if $S_{s.l.}$ tends to increase $\dot{\beta}$ algebraically, and negative otherwise.

A_s - The total area of the spin rocket exit nozzles.

A_p - The projection of A_s on a plane normal to the rocket axis.

A_e - The area of the main rocket exit nozzle (plus A_p if spin rockets are used).

m - The mass of the rocket (a function of time).

k - The radius of gyration about an axis through the center of gravity and normal to the rocket axis (a function of time).

k_o - The radius of gyration about the rocket axis (a function of time).

r_s - The distance from the rocket axis to the spin rocket exit nozzle.

r_o - The radius of gyration of the plane surface of the exit nozzle of the main rocket motor about the rocket axis.

With these definitions we define the following auxiliary coefficients for the equations of motion.

$$p = C_p - C_g \tag{1}$$

$$\xi_r = \sin\theta\cdot\cos\alpha \;;\; \eta_r = \sin\theta\cdot\sin\alpha \;;\; \zeta_r = \cos\theta \tag{2}$$

$$v_x = \dot{x} + u_x - p(\dot{\theta}\cdot\cos\theta\cdot\cos\alpha - \dot{\alpha}\cdot\sin\theta\cdot\sin\alpha) - \dot{p}\cdot\xi_r \tag{3}$$

$$v_y = \dot{y} + u_y - p(\dot{\theta}\cdot\cos\theta\cdot\sin\alpha + \dot{\alpha}\cdot\sin\theta\cdot\cos\alpha) - \dot{p}\cdot\eta_r$$

$$v_z = \dot{z} + u_z + p\cdot\dot{\theta}\cdot\sin\theta - \dot{p}\cdot\zeta_r$$

$$v_a = (v_x^2 + v_y^2 + v_z^2)^{\frac{1}{2}} \tag{4}$$

$$\xi_a = v_x/v_a ;\quad \eta_a = v_y/v_a ;\quad \zeta_a = v_z/v_a \tag{5}$$

$$\cos\delta = \xi_a\xi_r + \eta_a\eta_r + \zeta_a\zeta_r \tag{6}$$

$$\sin\epsilon = \omega^2 R_o \cdot \sin\lambda \cdot \cos\lambda / g_o \tag{7}$$

$$A_r = C_L \rho d^2 v_a{}^2 \mathrm{sign}(\cos\delta)/m \tag{8}$$

$$A_a = - C_D \rho d^2 v_a{}^2/m - A_r \cos\delta \tag{9}$$

$$A_m = - A_a/mk^2 \tag{10}$$

$$A_f = S/mk_o{}^2 \tag{11}$$

$$B_t = [T_{s.l.} + (P_{s.l.} - P_a)\, A_e]\, /m \tag{12}$$

$$B_j = [\dot{m}\,(h - C_g)^2 - 2m\,(h - C_g)\,\dot{C}_g - \dot{m}k^2 - 2mk\dot{k}]\,/mk^2 \tag{13}$$

$$B_r = [\dot{m}r_o{}^2 + 2mr_o\dot{r}_o - \dot{m}k_c{}^2 - 2mk_o\dot{k}_o]\,/mk_o{}^2 \tag{14}$$

$$B_s = s[S_{s.l.} + (P_{s.l.} - P_a)(A_s{}^2 - A_p{}^2)^{\frac{1}{2}} \cdot r_s]\,/mk_o{}^2 \tag{15}$$

$$C_w = \frac{- g_o R_o{}^2 \sin(\lambda + \epsilon)}{\sin\lambda[x^2 + (y - R_o\sin\epsilon)^2 + (z + R_o\cos\epsilon)^2]^{3/2}} \tag{16}$$

$$C_x = \omega[2\cdot\dot{y}\cdot\sin(\lambda + \epsilon) - 2\cdot\dot{z}\cdot\cos(\lambda + \epsilon) + \omega\cdot x] \tag{17}$$

$$C_y = C_w R_o \sin\epsilon - \sin(\lambda+\epsilon)[\omega\{2\dot{x}+\omega[z\cdot\cos(\lambda+\epsilon)-y\cdot\sin(\lambda+\epsilon)+R_o\cos\lambda]\}] \tag{18}$$

$$C_z = C_w R_o \cos\epsilon + \cos(\lambda+\epsilon)[\omega\{2\dot{x}+\omega[z\cdot\cos(\lambda+\epsilon)-y\cdot\sin(\lambda+\epsilon)+R_o\cos\lambda]\}] \tag{19}$$

$$C_n = k_o{}^2/k^2 \tag{20}$$

In equation (6), δ is the angle of attack, defined as the angle between the positive rocket axis and the velocity vector (relative to the surrounding air) of the center of pressure. In equation (7), ϵ is the small angle between the z axis (as measured by a plumb bob) and the pure gravitational force (i.e., with the centrifugal force due to the rotation of the earth removed).

Using the above definitions and auxiliary equations, the conservation laws of linear momentum and angular momentum give the desired equations of motion:

$$\ddot{x} = A_a\xi_a + A_r\xi_r + B_t\xi_r + C_w x + C_x \tag{21}$$

$$\ddot{y} = A_a\eta_a + A_r\eta_r + B_t\eta_r + C_w y + C_y \tag{22}$$

$$\ddot{z} = A_a\zeta_a + A_r\zeta_r + B_t\zeta_r + C_w z + C_z \tag{23}$$

$$\ddot{\theta} = A_m(\xi_a\cos\theta\cdot\cos\alpha + \eta_a\cos\theta\cdot\sin\alpha - \zeta_a\sin\theta) + B_j\dot{\theta} + (\dot{\alpha}\cdot\cos\theta + C_n\dot{\beta})\,\dot{\alpha}\cdot\sin\theta \tag{24}$$

$$\ddot{\alpha} = A_m(\eta_a \cos\alpha - \xi_a \sin\alpha)/\sin\theta + B_j\dot{\alpha} - (2\dot{\alpha}\cdot\cos\theta + C_n\dot{\beta})\dot{\theta}/\sin\theta \qquad (25)$$

$$\ddot{\beta} = A_f + B_r\dot{\beta} + B_s \qquad (26)$$

In these equations, the 'A' terms are non-zero only in the atmosphere, the 'B' terms are non-zero only during burning phases, but the 'C' terms, representing the effects of gravity, Coriolis force, centripetal force due to the rotation of the earth, and the gyroscopic effects, are never zero.

Most of the rocket data and atmospheric data will usually be given in table form. The values of C_p, C_L, C_D, and S are usually given as functions of Mach number; the values of C_g, m, $\dot{m}$, k_o, k, $T_{s.l.}$, and $S_{s.l.}$, as functions of time; and the values of u_x, u_y, u_z, ρ, P_a, and the velocity of sound, as functions of height. All of these quantities must be stored within the computer and, if given in table form, some interpolation routine used to find their values at specific points along the trajectory.

For many applications, these equations are more elaborate than is necessary. In these cases, some simplification is possible. For example, if one has a rocket which spins very slowly or not at all, the effects of spin can be removed by deleting equations (11), (14), (15), (20), and (26) and the β terms of (24) and (25). In most cases, the approximation, $\epsilon = 0$, instead of equation (7) does not alter results appreciably. The main reason for making such approximations is to decrease the time and cost of computation.

A word of caution should be given concerning these equations. They can accurately compute the hypothetical trajectory which will approximate an actual trajectory only to the extent that the atmospheric and rocket data are exact and the rocket itself is perfectly aligned.

III. THE USE OF PERTURBATION EQUATIONS

In many applications of the equations of motion one will find that he is interested in calculating many hypothetical trajectories which deviate only slightly from one hypothetical trajectory already calculated. This sort of thing will occur when making a statistical study using Monte-Carlo techniques, for example.

This will occur also when one is interested in computing a new trajectory which deviates from a previously calculated one because of a small change in one or more of the atmospheric or rocket parameters. In these situations, the use of perturbation equations is suggested. Perturbation equations describe the changes from the original trajectory rather than define a new hypothetical trajectory. The reason for doing this is two-fold. First of all, the perturbation equations can usually be made simpler than the original equations by making certain first-order approximations. Secondly, some of the calculations which have been made in computing the original trajectory can be used in the perturbation equations.

The specific perturbation equations to be used depend upon the purpose for which they are intended. The derivation of the desired perturbation equations from the equations of motion is usually done by using ordinary differential calculus and by making certain approximations. The perturbation equations are accurate only for small deviations from the original hypothetical trajectory. What constitutes a small deviation is governed by the problem.

For explanation of the general ideas in perturbation theory the reader is referred to (5). More elaborate sets of perturbation equations can be derived which use second-order (or higher) variations; i.e., one uses not first but second-order approximations. To do this, however, yields much more complex perturbation equations. The author feels that linear approximations offer the most efficient equations for the purpose. For the derivation of perturbation equations which describe the deviation of a trajectory due to changes in the wind profile see (6).

In general, perturbation equations will accelerate computation time greatly. However, one should realize that they are inherently not exact, but only approximate. Before using this technique one must test the accuracy of the perturbation equations. This is done by comparing results with the results one would get if he used only the equations of motion for the entire problem. Of course, the perturbation equations are, by their nature, more accurate for small perturbations.

IV. THE NUMERICAL INTEGRATION

Since the equations of motion of Section II and the perturbation equations of Section III are simultaneous, non-linear, differential equations and since the coefficients used in these equations are not constant (and in many cases not even easily expressed analytically) one is required to integrate them by strictly numerical methods. Many numerical integrating techniques can be found in the literature. Each of these has its advantages and disadvantages. For applications to equations of motion it is felt that the Runge-Kutta integration scheme is best.

There are several reasons for choosing the Runge-Kutta technique. First, the six equations of motion, being somewhat different in form, require an integration technique which is flexible enough to handle all of them simultaneously. Second, the size of the integration interval, Δt, to be used need not be constant. For accuracy and time consideration, Δt should be small during the initial part of the trajectory, while during the latter part it may be increased. The Runge-Kutta technique lends itself nicely to this problem, a feature that most other integration schemes do not have. Third, since any numerical scheme will have inherent errors, some measures should be taken for controlling them. Again the Runge-Kutta scheme seems best. For details of this technique see (7) and (8).

For a mathematical model to be as effective as possible, one should make provisions in the computer program for changing rocket parameters at a number of arbitrary points along the trajectory. This

is necessary if he wants to calculate hypothetical trajectories for rockets which have one or more boosters, or for various reasons might have some of their parameters changed instantaneously during flight. The ultimate goal is to have a program which, in one operation, will yield the entire trajectory. Also, the program should be general enough for application to various types of rockets. Such a program has been written at USASMSA which has all of the above features (except the features of spin). More will be said about this program in the summary.

V. PRACTICAL APPLICATION

Because of the complexity of the equations of motion, it seems that any statistical study made on the trajectory of a rocket must, in the final analysis, involve Monte-Carlo techniques. Roughly, this means that one must compute a number of trajectories, the cross section of which would represent the probable population of trajectories. This immediately suggests the use of perturbation equations.

One example is the calculation of the statistical parameters involved in the results of (9), which gives the theory for computing the dispersion pattern of a given trajectory parameter as a function dispersion of the wind profile.

As a practical application of the equations of motion and the perturbation equations, we show how they are used in the support of the actual firing of a research rocket. We first make the assumption that the rocket is perfectly aligned and that all of the rocket and atmospheric data are exact. Because of the turbulence of the wind at various points along the expected trajectory, it is exceedingly difficult to predict what the wind will be at a specific point along the trajectory and at the precise time that the rocket will be at that point. That is, the turbulence causes a certain amount of unreliability in the wind prediction, and hence in the impact prediction. We wish to compute not only the most probable impact point (MPIP), but also the size and shape of the area within which the rocket has, say, a 99 percent probability of falling. This first MPIP is computed from the equations of motion (21)-(26) using the most probable wind profile (MPWP) which has been predicted at some time prior to the proposed firing time. From the turbulence data, one can compute many wind profiles, the aggregate of which will represent the probable distribution of the predicted wind profile about the MPWP. Using the perturbation equations, which describe displacement due to a change in the wind, with each of the wind profiles in the aggregate, one gets a distribution of the rocket impact point about the MPIP. But all of this data collection and calculation takes a certain amount of time, during which the wind will have changed. If one now measures the wind and makes a new prediction on the MPWP, he can use the perturbation equations to get a new MPIP. With the assumption that the turbulence has remained the same, he can simply translate the impact dispersion pattern to the new MPIP.

This now gives the range safety officer a complete picture of the variability of the trajectory due to the variability of the wind only. The same word of caution that was given at the end of Section II applies here. That is, the results, being hypothetical, are accurate only if the atmospheric and rocket data are accurate and the rocket is correctly aligned.

VI. SUMMARY

This model has been programmed at USASMSA using a fourth-order Runge-Kutta integration technique with all of the features indicated in Section IV. It was written for the Philco 2000 Computer and required about 500 man hours to write. Most of this time was spent in making the program general enough to compute in one operation the hypothetical trajectory of any unguided rocket (with range less than 300 miles) with any combination of boosters. All that is necessary as input is the pertinent rocket and atmospheric data.

As mentioned in Section IV, any numerical integration scheme has inherent errors. In the program at USASMSA, these errors can be made as small as desired, at the expense of computer time. To give an example of computation speeds, for a typical Aerobee rocket which has a time of flight of about 7 minutes, the error (due only to the numerical integration inaccuracies) at the impact point can be made to be no greater than 100 feet with a computing time of less than 20 minutes. With a proposed modification to the computer this time should be reduced to less than 5 minutes. It is anticipated that perturbation equations for a typical Aerobee flight can be integrated in less than 2 seconds.

The model together with the computer program has many applications at USASMSA. One of these was discussed in the previous section. It could also be used for preliminary studies to determine the feasibility of firing a given rocket from a given range.

Another application is to test various theories concerning rocket flight. One of these is the effect of certain wind profiles on the stability of the rocket. A closely related study is to test the effects of so-called acoustical turbulence suggested in (10). In testing such theories a mathematical model is much more economical than an actual rocket firing; in fact, in many cases (for example, the two just mentioned) this is the only way.

In general, regardless of whether the mathematical model used is the one proposed in this paper or some other, its use will usually result in an appreciable savings in time and cost. It can be used in support of rocket design and firing and in testing certain theories of rocket flight.

REFERENCES

(1) Lewis, J. V., "The Effect of Wind and Rotation of the Earth on Unguided Rockets," Ballistic Research Laboratories Report No. 685, Aberdeen Proving Ground, Maryland, 1949.

(2) Daw, H. A., "A Wind Weighting Theory for Sounding Rockets Derivable from the Rocket Equations of Motion," Physical Science Laboratory, New Mexico State University, University Park, New Mexico, 1958.

(3) Walter, E. L., "Six-Variable Ballistic Model for a Rocket," U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, 1962.

(4) Rosser, Newton, Gross, "Mathematical Theory of Rocket Flight," McGraw-Hill, New York, 1947.

(5) Forsyth, A. R., "Calculus of Variations," Dover, New York, 1927.

(6) Walter, E. L., "Wind Effects for the Six-Variable Ballistic Model from Perturbation Theory," U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, 1962.

(7) Hildebrand, F. B., "Introduction to Numerical Analysis," McGraw-Hill, New York, 1956.

(8) Gill, S., "A Process for the Step-by-Step Integration of Differential Equations in an Automatic Digital Computer Machine," Proc. Cambridge Phil. Soc., Vol. 47, pp 96-108, 1951.

(9) Beer, F. P., Sarubbi, R. G., "On Influence of Atmospheric Gustiness on the Trajectory of Free Rockets," Lehigh University, Bethlehem, Pa., (Under a Research grant from U. S. Army Signal Corps), 1961.

(10) Webb, W. L., Jenkins, K. R., "The Sonic Structure of the Mesophere," Journal of Acoustical Society of America, Vol. 34, No. 2, Feb., 1962.

AN APPARATUS FOR MEASURING THE BULK MODULUS OF SOLID PROPELLANTS

NEAL C. WOGSLAND
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION

The mechanical behavior of the propellant is of considerable importance in the design of a solid-propellant rocket because the propellant must be sufficiently stiff to minimize creep during storage and yet must not be so brittle that it may crack under the pressures at which it burns. Some solid propellants are homogeneous compounds while others are composite mixtures of crystalline oxidizers and other additives in a small amount of binder. Both types of solid propellant, however, are predominantly viscoelastic in their mechanical behavior.

The basic methods for solving problems in stress analysis are well established for elastic materials. The classical theory of elasticity is limited almost entirely to situations in which the strains are small. The strains set up within an elastic material are linear functions of the stresses under all conditions of loading, provided the elastic limit is not exceeded; that is

$$\frac{\text{stress}}{\text{strain}} = \text{constant (modulus of elasticity).} \qquad (1)$$

When the strains are small, only two elastic constants are required to describe the mechanical behavior of a material that is homogeneous and isotropic. If, for example, the shear modulus (modulus of rigidity), G, and the bulk modulus, K, are determined, then Young's modulus, E, and Poisson's ratio, ν, can be calculated from the relations

$$E = \frac{9KG}{3K + G} \quad \text{and} \quad \nu = \frac{3K - 2G}{6K + 2G} \qquad (2, 3)$$

which have been given by Southwell (1) a[illegible]thers. Any of these moduli may be replaced by their correspo[illegible]g reciprocals, 1/J, 1/B,

and 1/D, where J is shear compliance, B is bulk compliance (compressibility), and D is tensile compliance. The shear and bulk moduli may be regarded as the fundamental elastic constants from a physical point of view because the former measures the resistance of a material to change of form unaccompanied by change of volume and the latter measures its resistance to change of volume unaccompanied by change of form.

The stress analysis of viscoelastic materials is somewhat more complicated, not only because these materials are time-dependent, but also because they deviate measurably from the behavior predicted by the Boltzmann superposition principle (2, 3) on which the theory of linear viscoelasticity is based. For small strains, however, this deviation can be neglected for many materials and calculations can be based on linear viscoelastic behavior. In the final report of the Committee on Nomenclature of the Society of Rheology, Leaderman (4) stated that the relations between the complex moduli and compliances for dynamic viscoelastic behavior are the same as those existing between the corresponding moduli and compliances of classical elasticity theory.

Stress analysis problems involving viscoelastic materials frequently can be formulated in terms of Volterra integral equations if the moduli are known (5, 6). An alternate method sets up analogies using mechanical models consisting of suitable combinations of Hookean springs and Newtonian dashpots (7, 8). In current applications of linear viscoelastic theory, the distinction between moduli measured under adiabatic and isothermal conditions usually is ignored because this difference is in most cases negligible. Although this difference is entirely negligible in shear and the moduli seldom differ by more than a few percent in bulk, Ferry (3) has inferred on theoretical grounds that the adiabatic bulk modulus can exceed the isothermal by 20% or more for some soft polymeric solids.

At the Ballistic Research Laboratories, a study is being conducted to determine to what extent the theory of linear viscoelasticity can be applied to the stress analysis of solid propellants (9) To correlate theory with experimental results, apparatus are being devised to measure the bulk and shear moduli under various conditions. Preliminary study has indicated that the equilibrium bulk modulus is satisfactory for quasi-static problems such as the slow deformation of a propellant in storage. However, the adiabatic bulk modulus is better for approximate solutions of dynamic problems.

The apparatus described in this paper was built to measure the equilibrium bulk modulus of solid propellants and other viscoelastic materials under hydrostatic pressures of 0 to 2500 psi (gage). It was built after preliminary tests had indicated that such an apparatus was feasible (10). Very little information has been reported in the scientific literature concerning hydrostatic modulus or compliance testing in this pressure range. Bridgman (11) has

worked with several viscoelastic materials, but only at very high pressures. Sweeny and Bills (12) have made a few pressure-volume measurements on solid propellants. Milloway, Surland, and Skulte (13) have recently reported that the composite propellants used in their testing are sufficiently isotropic in hydrostatic compression that bulk modulus can be calculated from uniaxial displacement measurements.

APPARATUS AND EXPERIMENTAL PROCEDURE

This apparatus is designed to determine the bulk modulus of a solid by substituting the solid for some of the liquid in the test chamber, measuring the net compressibility of the solid plus the remaining liquid, and applying differential methods to separate their compressibilities. The intensifier principle is utilized to obtain high sensitivity in measurement of the volume changes produced by variation of the hydrostatic pressure. The test sample is placed directly in the high-pressure chamber of the intensifier to minimize the volume of fluid required and therefore minimize the influence of the fluid on the test data (Figure 1). A test chamber of approximately 1 cu. in. was selected so that small propellant samples (up to 1/2 in. in diameter by 3 in. in length) can be tested effectively.

The apparatus was designed for operation at pressures from 0 to 2500 psig because this range covers the pressures that occur during the burning of most solid-propellant rockets. However, the intensifier principle could be used for much higher pressures (or larger chamber volumes) with appropriate design. The piston seals are a modification of Bridgman's (11) unsupported area seal. The slipperiness of the teflon seals permits observation of pressure-volume data at pressures close to 0 psig even though the seals are preloaded to minimize leakage at these low pressures.

In addition to the pressure intensifier, the bulk modulus tester utilizes auxiliary equipment and instrumentation including a sight-glass system, hydraulic pumps, valves and fittings, a dial pressure gage, a pressure transducer, a strain indicator or recorder, a thermocouple, a microvolt amplifier, and a strip-chart recorder. A photograph of the system is shown in Figure 2.

The total volume of the test chamber and its connecting passageways was obtained by measurement and calculation. The relationship between the volume change in the chamber and the sight-glass reading also was calculated. The pressure transducer was calibrated in conjunction with the strain-indicator and a standard dead-weight tester. The chromel-alumel thermocouple was set up to measure the temperature change within the chamber. The temperature scale of the recorder chart was established from thermocouple data sheets since they provided sufficient accuracy for this test.

During the initial setup of the apparatus, a vacuum pump is used to evacuate each chamber as it is filled with hydraulic fluid.

Hercoflex 600 hydraulic fluid has been used with solid propellants at the BRL. Hercolube A or Dow-Corning DC-550 probably would be equally satisfactory. If necessary, the test sample may be given an impermeable coating to prevent penetration of the fluid into the pores or to prevent a possible chemical reaction. When a sample is prepared for testing, the test chamber is filled to overflowing before installing the closure so that no air will be entrapped. This precaution is essential because a major problem in an apparatus of this sort is the complete elimination of air in the test chamber. For critical tests, much of the air dissolved in the working fluid can be removed by boiling under vacuum prior to filling the system.

The apparatus was calibrated by measuring the pressure and volume changes for steel samples of different sizes. With the sample in the high-pressure chamber, pressure is applied to the low-pressure chamber by a hand pump. The pressure is intensified as it is transmitted to the smaller chamber by the double-faced piston, and as the piston moves, the excess fluid enters the sight glass from the middle chamber. Measurement of the volume displaced by the low-pressure end of the double piston provides a very sensitive measurement of the volume change in the high-pressure chamber. During isothermal testing, an interval of approximately 10 minutes is allowed prior to recording the data for each pressure so that temperature equilibrium will be reached. The second pump is used to return the piston to its starting position after each test is completed.

To obtain adiabatic data, the system is quickly pressurized from 0 psig to a desired pressure and the pressure-volume-temperature data is recorded immediately. The pressure is then dropped to 0 psig and the system is allowed to return to equilibrium temperature. The cycle of operation is repeated for each additional pressure that is needed.

ANALYSIS AND DISCUSSION

As previously stated, this apparatus was designed to measure the equilibrium bulk modulus of solid propellants and other viscoelastic materials. The bulk modulus is the ratio of the volumetric stress (hydrostatic pressure) to the volumetric strain (volume change per unit volume),

$$K = - \frac{\Delta P}{\Delta V/V} \cdot \tag{4}$$

The negative sign is introduced because hydrostatic tension is normally considered positive and the application of a hydrostatic pressure produces a decrease in volume. In the direct measurement of the bulk modulus of a solid by immersion in a liquid, differencing techniques are applied to correct for the compressibility of the fluid and for the distortion of the test chamber. The initial volume of the chamber is equal to the sum of the sample and fluid volumes,

$$V_c = V_s + V_f \ . \tag{5}$$

When pressure is applied by movement of the piston, the new chamber volume is

$$V_c + \Delta V_c = V_s + \Delta V_s + V_f + \Delta V_f = V_c + C'\Delta P - \Delta s \pi r^2 \ , \tag{6}$$

where C' is a constant that accounts for the pressure effects on the chamber itself (expansion of chamber, pressure gage, and connecting passageways; compression of seals), Δs is piston travel, and r is the radius of the piston. Second order effects are small in the pressure range under consideration, and their effect on the data are minimized by the differencing technique described herein. The volume change per unit pressure is

$$\frac{\Delta V_c}{\Delta P} = \frac{\Delta V_s}{\Delta P} + \frac{\Delta V_f}{\Delta P} = - \frac{C''\Delta h}{\Delta P} + C' \ , \tag{7}$$

where Δh is the change in the height of the sight glass column and C" accounts for the magnification of the piston motion by the sight glass system.

Combining Eq. (4) and Eq. (7), the bulk modulus of a test sample may be written

$$K_s = - \frac{V_s}{\Delta V_s/\Delta P} = \frac{V_s}{(C''\Delta h/\Delta P) - (V_f/K_f) - C'} \ . \tag{8}$$

The bulk modulus of the sample is determined from the measurements of the sight glass reading, Δh, and the applied pressure, ΔP, after the other items have been established. The initial volume of the chamber, V_c, is determined by measurement and calculation. Measurement of the initial sample volume, V_s, then yields the initial volume of fluid, V_f. The sight glass magnification constant, C", is calculated from the sight glass and piston diameters. The bulk modulus of the fluid, K_f, and the chamber expansion constant, C', are determined during calibration of the apparatus.

The apparatus is calibrated by using various sizes of test samples with a known bulk modulus. The reference samples that were used were made from cold-rolled steel, which has a bulk modulus of approximately 23.1×10^6 psi (14), nearly 100 times that of the transmitting fluid. From Eq. (8), it is seen that the equation

$$\frac{C''\Delta h}{\Delta P} - \frac{V_s}{K_s} = \frac{V_f}{K_f} + C' \tag{9}$$

is an equation of the linear form y = mx + b. On a plot of V_f vs $\Delta V/\Delta P$, the above equation represents a straight line with a slope of $1/K_f$ and an intercept on the $\Delta V/\Delta P$ axis of C'. The measurement of Δh

and ΔP for two different fluid-sample combinations is sufficient to determine the slope and the intercept. In practice, seven combinations were used and the method of least squares was applied to obtain the best average slope and intercept.

During the tests of samples of various materials, the minimum number of observations per test was six and the method of least squares was again applied to obtain the best average slope $\Delta h/\Delta P$. The bulk modulus was then calculated from Eq. (8). Equilibrium bulk moduli of several viscoelastic materials and the transmitting fluid are tabulated in Table I for two pressure ranges, 0-2500 psig and 500-2500 psig.

TABLE I. EQUILIBRIUM BULK MODULI, IN PSI, OF SEVERAL VISCOELASTIC MATERIALS AT $77 \pm 1^{\circ}F$

Material	Pressure Range, psi (gage) 0 - 2500	500 - 2500
Hercoflex 600 Fluid	249,000	252,000
Polystyrene (commercial rod)	472,000	476,000
Teflon (commercial rod)	363,000	367,000
Polyurethane (pluracol base)	258,000	267,000
Propellant No. 1 (cast double base)	432,000	438,000
Propellant No. 2 (composite double base)	583,000	621,000
Propellant No. 3 (composite)	834,000	882,000

The bulk modulus of each material is quite linear between pressures of 0 and 2500 psig when there are no voids. However, porous materials show large differences between their average moduli over the two ranges tabulated here because the voids undergo considerable compression at low pressures. With soft composite propellants, the effect of voids on the compressibility can be ignored above approximately 100 psig. The tabulated moduli for the propellants are taken from the second compression cycle for each material because the initial cycles were very non-linear while the voids were being compacted. The stress-strain relationship observed with teflon, a fairly dense material, is illustrated in Figure 3. In contrast, Figure 4 shows a composite double-base propellant with an initial porosity of approximately 0.33%. The propellant is compacted on the first test cycle, resulting in a porosity of less than 0.05% on the second run, which was made several hours later.

As indicated in the preceding paragraph, the porosity of a soft viscoelastic material can be determined with this apparatus. The intercept of the observed stress-strain curve with the strain-axis at zero applied stress provides a measure of the strain resulting from closure of voids. The porosity of the material is the ratio of this strain to the initial sample volume. Porosities of 0.01% can be measured by this procedure. The effect of voids on the compressibility data can be observed closely at low pressures because the

slippery teflon piston seals permit the piston to creep in the cylinder at applied loads considerably below the preload on the seals.

Actually, the volumetric stress-strain ratio for a material is not linear because its resistance to compression increases as its volume is reduced. Bridgman (11) and others have verified through extensive testing that, except for a few rare cases, the compressibility decreases with increasing pressure, as would be expected. This can be observed on the curves of Figure 3, where the data points are connected by a smooth curve, and Figure 4, where the points are represented by straight lines. Further analysis of the experimental data indicates that the tangent bulk modulus increases linearly with pressure in the manner predicted by the Tait equation of state for liquids (15),

$$- \frac{dP}{dV} = mP + b \ , \tag{10}$$

where m is a constant, b is a function of temperature only, and $-dP/dV = K/V$ [Eq. (4)]. The constants are easily established by graphical or analytical methods. Isothermal data from tests of the fluid, a teflon sample, and a propellant sample indicate that the Tait equation provides an accurate fit to the experimental data after the voids, if any, are closed. This equation should also be useful in correlating isothermal and adiabatic data.

The precision of measurement of this apparatus is such that changes of .00001 cu. in. in the volume of the chamber are readily detected and changes to .01 cu. in. can be measured with the present sight glass. These measurements compare with the initial chamber volume of .942 cu. in. and a maximum sample size of .57 cu. in. Larger volume changes could be measured with a glass of larger diameter, but at a sacrifice in precision. The use of two sight glasses in tandem would permit testing of very porous materials, using the larger glass for porosity data and then switching to the precision glass for modulus data.

The accuracy of the bulk moduli measured with this apparatus has not been established because a search of the literature has not revealed a suitable reference standard for this low pressure range. Based on a 95% confidence level, the bulk modulus of a test sample should be repeatable within 2%. However, this does not account for the deviations during calibration and systematic errors in measuring the dimensions of the sight glass and test chamber. There also is the problem of measuring the initial volumes of test samples having irregular shapes. A sample of an unknown aluminum alloy, which should have a bulk modulus close to 10×10^6 psi (about 30 times that of the transmitting fluid), was tested while running the calibration tests with the steel samples. Its modulus appeared to be between 9 and 12×10^6 psi. This implies accuracies within ±2% for viscoelastic materials, which is questionable.

The hydraulic fluid and viscoelastic samples are very sensitive to temperature changes. When a sample of maximum size (61% of chamber volume) is under test, the thermal expansion of the fluid (39%) for a temperature rise of 1°F is as great as the volume reduction of the aluminum sample at 2500 psig. The thermal expansion of the test sample may increase the total expansion by another 50%. Therefore, extremely close control over temperature must be exercised.

Although this apparatus was designed specifically for equilibrium (isothermal) testing, the measurement of temperature change inside the test chamber in conjunction with the P-V data permits determination of the adiabatic bulk modulus of the transmitting fluid and estimation of the adiabatic bulk modulus of the test sample. The modulus of the fluid is separated from the equivalent modulus correction for the apparatus through differential analysis of the calibration test data, as has been done for the isothermal testing. This same technique is then applied to estimate the adiabatic modulus of the test sample. Since these results are quite inaccurate, adiabatic testing has been limited to exploratory runs. A larger apparatus, in which a thermocouple is imbedded in the sample and the fluid volume is minimized, would be more desirable for adiabatic measurements.

A major problem in this apparatus was complete elimination of air in the pressure chamber. A preliminary apparatus (10) assembled from standard high-pressure laboratory equipment was only partially successful because of this defect. The new apparatus was designed to permit evacuation of the test chamber, and the other chambers, from one side while filling from the other. Each individual test was set up with great care to insure the reliability of the test data.

This compressibility apparatus is relatively safe since it avoids the danger associated with the large kinetic energies of gas-filled pressure bombs. If any sudden surge of pressure should occur, it would cause the pressure transducer to fail without danger of flying metal fragments. However, with minor changes, the system could be set up for remote operation.

SUMMARY AND CONCLUSIONS

This apparatus provides a measure of the equilibrium bulk modulus and compressibility of solid propellants and other viscoelastic materials at pressures between 0 and 2500 psig. The test sample is placed in the high-pressure chamber of a pressure intensifier. As hydrostatic pressure is applied, measurement of the volume displaced by the low-pressure end of the double piston provides high sensitivity in the measurement of the volume reduction in the high-pressure chamber. The measured isothermal modulus is sufficiently accurate for use in the quasi-static stress analysis of solid-propellant rockets. In addition, the adiabatic bulk modulus can be estimated from pressure-volume-temperature data. Porosities as small

as 0.01% also can be measured for soft viscoelastic materials and the rapid initial rise in the pressure-volume curve shows the effect of voids on the bulk modulus. Test results indicate that porous materials do not fully recover from the compacting effect of hydrostatic pressure, thus indicating that such a procedure might be used to increase the density of porous propellants.

This apparatus is quite satisfactory for the measurement of the bulk modulus of small samples (up to 1/2 in. in diameter by 3 in. in length) and it is especially applicable to the testing and screening of the small samples normally available with newly-synthesized experimental propellants. The principle weakness of this apparatus is the masking of the properties of the test sample by the transmitting fluid, which fills at least 39% of the test chamber, is 1 to 3 times as compressible as the sample, and has a relatively high coefficient of thermal expansion. However, the apparatus is safe and its operation is fairly rapid because it employs a continuous reading method. Where more material is available for test samples, a larger test chamber could be utilized. To provide for greater accuracy in both isothermal and adiabatic testing, the volume of fluid should be minimized and a thermocouple should be imbedded in the test sample.

REFERENCES

1. Southwell, R. V. An Introduction to the Theory of Elasticity for Engineers and Physicists. London: Oxford University Press, 1941.
2. Tobolsky, A. V. Properties and Structure of Polymers. New York: Wiley, 1960.
3. Ferry, J. D. Viscoelastic Properties of Polymers. New York: Wiley, 1961.
4. Leaderman, H. Proposed Nomenclature for Linear Viscoelastic Behavior. Transactions of the Society of Rheology, Vol. 1. New York: Interscience, 1957.
5. Lee, E. H., and Rogers, T. G. Solution of Viscoelastic Stress Analysis Problems Using Measured Creep or Relaxation Functions. Providence, R. I.: Brown University, Aug 1961.
6. Elder, A. S. Derivation of Equations for the Stresses and Strains in a Cylindrical Viscoelastic Case-Bonded Grain. Aberdeen Proving Ground: BRL MR 1359, 1961.
7. Bland, D. R. The Theory of Linear Viscoelasticity. New York: Pergamon Press, 1960.
8. Freudenthal, A. M., and Geiringer, H. The Mathematical Theories of the Inelastic Continuum. Encyclopedia of Physics, Vol. VI, Elasticity and Plasticity. Berlin: Springer-Verlag, 1958.
9. Elder, A. S. Stress Function Theory for Linearly Viscoelastic Solids. Aberdeen Proving Ground: BRL MR 1282, 1960.
10. Mears, J. W., and Wogsland, N. C. An Apparatus for Measuring Bulk Modulus of Viscoelastic Materials Under Hydrostatic Compression. Aberdeen Proving Ground: BRL TN 1321, 1960.
11. Bridgman, P. W. The Physics of High Pressure. London: Bell, 1949.

12. Sweeny, K. H., and Bills, K. W., Jr. Poisson's Ratio Determination; Compressibility Measurement. Bulletin of the Seventeenth Meeting of the JANAF Panel on Physical Properties of Solid Propellants: 111-112. Silver Spring, Md.: Johns Hopkins University, SPIA PP-11, 1958 (Confidential).

13. Milloway, W. T., Surland, C. C., and Skulte, I. The Effect of Initial Voids on the Bulk Modulus and Void Formation on Uniaxial Extension. 20th Meeting Bulletin, JANAF-ARPA-NASA Panel on Physical Properties of Solid Propellants, Vol. 1. Silver Spring, Md.: Johns Hopkins University, SPIA PP-14u, 1961.

14. Vose, R. W. Mechanical Properties of Materials. <u>Mechanical Engineers' Handbook, Ed. by L. S. Marks, 4th Ed.</u> New York: McGraw-Hill, 1941.

15. Hirschfelder, J. O., Curtiss, C. F., and Bird, R. B. <u>Molecular Theory of Gases and Liquids</u>. New York: Wiley, 1954.

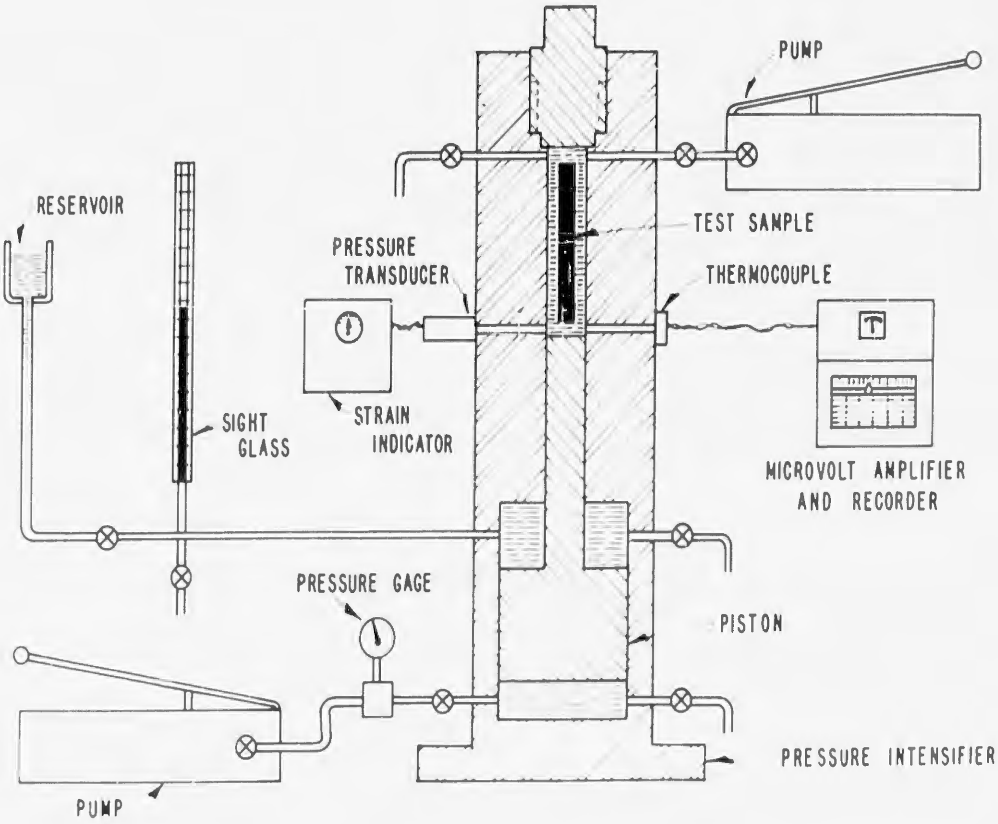


Figure 1. Schematic Drawing of Bulk Modulus Apparatus

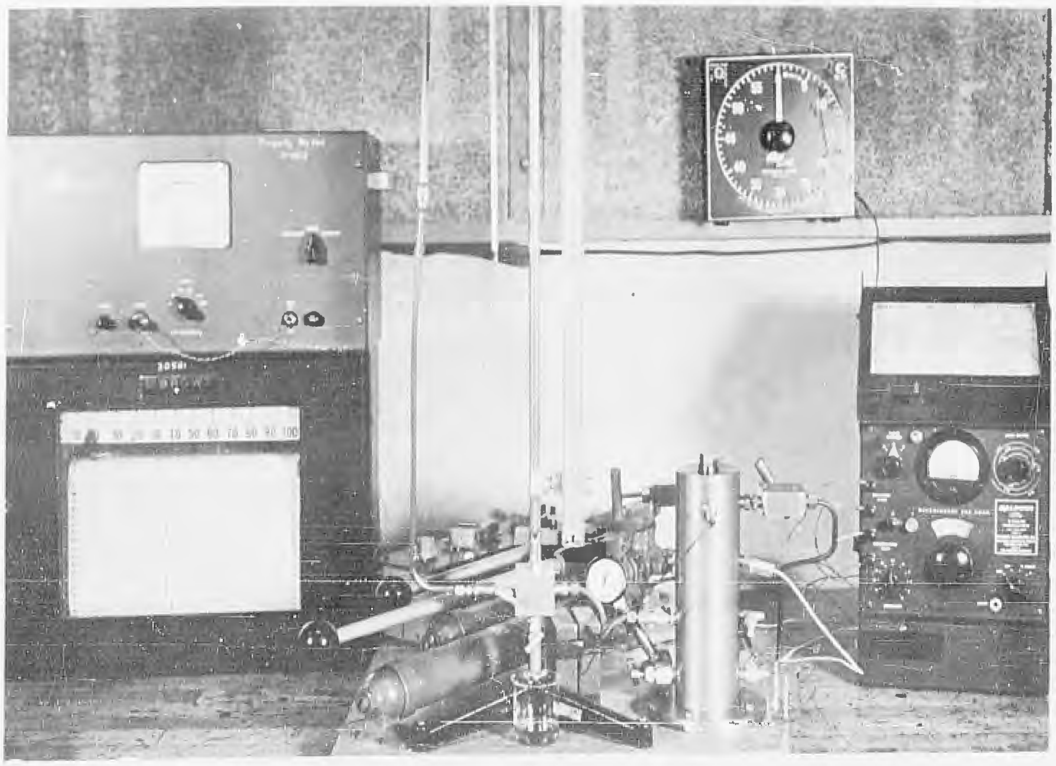


Figure 2. Bulk Modulus Apparatus

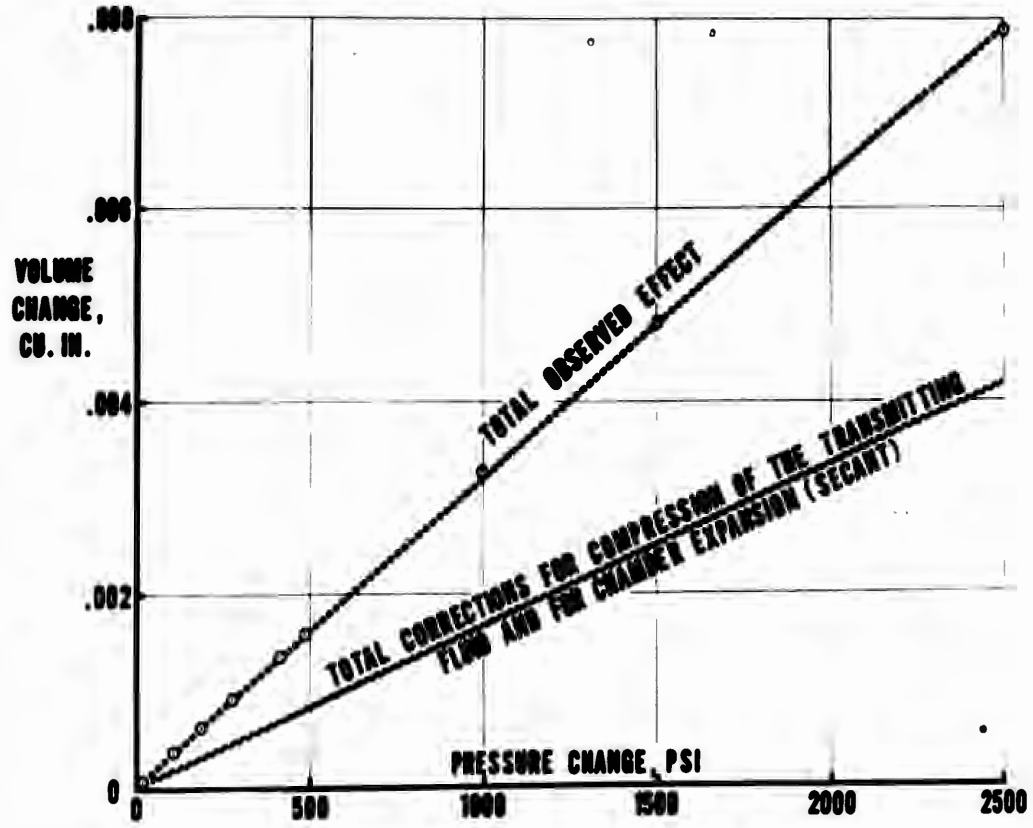


Figure 3. Pressure-Volume Curve for a Teflon Sample

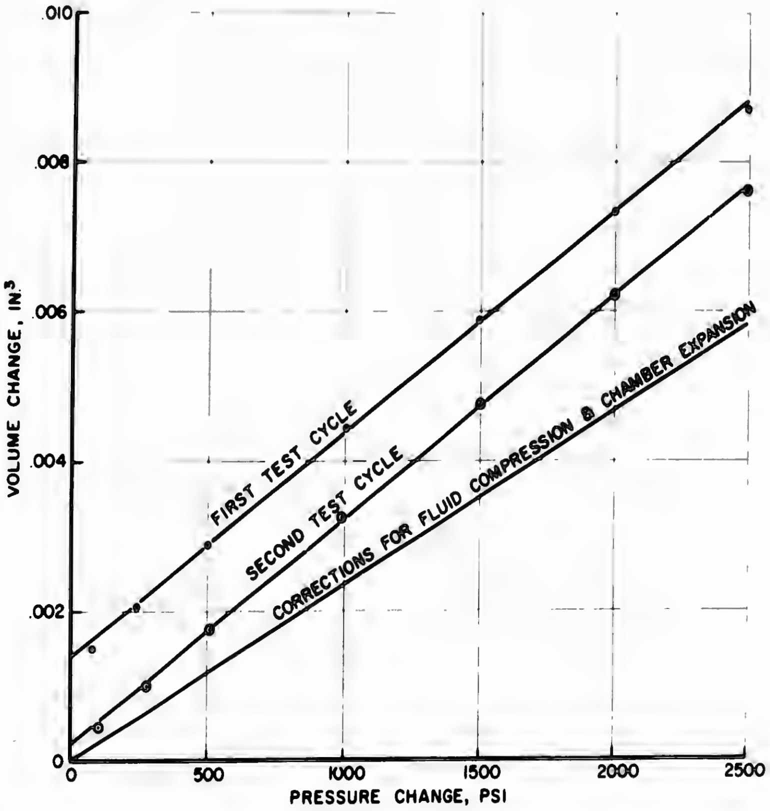


Figure 4. Pressure-Volume Curves of Propellant of 1/3% Initial Porosity

A FLUID AMPLIFIER HEART PUMP

K. E. WOODWARD, T. G. BARILA, D. E. NUNN,
G. MON, and H. H. STRAUB
DIAMOND ORDNANCE FUZE LABORATORIES
WASHINGTON, D. C.

The Problem

The advent of open heart surgery has presented to the medical profession the opportunity of repairing damaged or diseased hearts that otherwise might soon fail completely. Many devices are involved in this type of surgery. One important apparatus is a pump that can assume the heart's role while the heart is being repaired. Existing pumps are fairly complicated both in design and in the methods of control.

Fluid amplification (ref 1, p 22), recently invented by a team of Army scientists, offers the possibility of extreme design and control simplification for pulsatile types of extracorporeal heart pumps and a more accurate duplication of the heart's physiological pumping functions. Because all moving control parts and electronics can be eliminated, reliability can be measurably improved and production costs can be significantly reduced. These unusual qualifications recommended such a pump to the Army Medical Service for field use.

Application of the pump might be expanded to include the support of failing circulations, and hearts damaged or diseased, which with support could recover.

Consequently the problem was to design an extracorporeal heart pump powered and controlled by fluid amplification principles. This paper describes the development effort to date.

Design Requirements

The requirements for any extracorporeal heart pump may be broken down into three major categories. The pump must:

a. Duplicate the heart's essential pumping functions;
b. Possess adequate reliability and life;
c. Be appropriately packaged.

Each of these will be considered in detail.

a. Physiological Requirements

The pump must obviously possess enough of the essential functional pumping characteristics of the human heart to sustain life. Furthermore, irreversible blood damage can not be tolerated either during or after the pumping run. From an examination of the literature and existing pump designs, it appears that a single set of functional requirements for extracorporeal pumps in general has not been established. The authors in developing the fluid amplifier pump considered the following heart functions important:

(1) Output Pressures and Flows

Certainly the pump must be capable of adequate perfusion of the subject. Moreover it seems desirable to have a single pump satisfy the needs of a relatively large range of subjects regardless of age or size. The pertinent cardiovascular pump parameters (pressures, flows, age, size, pulse and activity level) for the human male from age 1 through 80 are summarized in Figure 1. The ordinate is a cardiac flow index relating mean blood flow rate to total body surface area. Since the parameters of Figure 1 are collated from several medical texts, the information lacks a high degree of statistical significance. However, it does establish the general output requirements for the pump. For operating room application, only basal or resting conditions are of concern.

(2) Load-Matching Capabilities

Reference 2 suggests that the human heart may recognize the flow resistances of the cardiovascular circulations and then exert just enough myocardial (muscle) force per stroke to achieve the necessary blood flows. Excess force causes excessive blood damage. An index of such trauma is the increase in red cell destruction. This index is called hemolysis. To control hemolysis and other blood damage the pump should be designed to propel blood with minimal force application.

(3) Pulsatile Blood Flows

Disagreement exists among cardiovascular experts relative to the need for pulsatile blood flows for extracorporeal pumps. The technically easier constructed nonpulsatile pumps are currently most popular but both pulsing and steady-flow pumps presently function successfully for short periods - one to two

hours. It may be, however, that long-term perfusion requires pulsed flows. Reference 2 suggests that resonant principles may operate in the cardiovascular system and may, among other things, enhance body tissue perfusion. If so, the pumping system (including the catheter) must then be capable of duplicating the subject's normal pulse to exploit the elastic responses of the circulatory systems.

(4) Vasoconstrictive Considerations

The heart is a relatively constant myocardial force pump because of vessel distensibility and the vasoconstrictive characteristics of the cardiovascular system. Flow resistances are made to vary autonomously by the nervous and endocrine systems to suit the physical and emotional needs of the individual. To minimize the pump forces propelling the blood and to allow the pump to regulate its output in response to vasoconstrictive changes in flow resistance, output flows must decrease with increased flow resistance and vice versa. To preclude packing or depletion of the two circulation systems, the output flow must also be made to increase with increased filling pressures, as in the case of the human heart.

(5) Other Functional Requirements

Filling of the human heart ventricle is achieved mainly by the difference between atrial and ventricular pressures during the period the difference is maximal. However, atrial contraction plays a small but significant part in the filling process (ref 3, p 244) and serves to increase the flow rate of the heart by both increasing ventricular distention and compensating for valve resistance. For these reasons and if cannulae* resistances are significant, it might be desirable to cause the pump to suck slightly in diastole. Additionally if the pump is to be used as an augmentation device for failing heart patients, the pump should be capable of synchronizing its pulses with that of the diastolic phase of the heart. By pumping blood during diastole, both peak myocardial forces and the work of the heart are diminished.

The pump should also cause as little blood damage as that caused by presently available commercial pumps. Blood damage arises not only from the use of unnecessarily large pump pressures, referred to above, but also from a variety of other phenomena, such as turbulence and shear, that are influenced by the mechanical design of the pump.

b. Reliability and Life Considerations

High reliability is an obvious necessity. It should be measurable and predictable and as independent as possible

*Cannulae are small tubes inserted into the body cavity through which blood passes from and to the pump.

of operating conditions, the type of personnel responsible for the pump, and the kind of power source used. In the event of power failure, the pump should be designed to operate on stored energy. The pump should require minimum maintenance.

c. Packaging

The following design attributes would be generally desirable for the pump. It should be:

(1) Simple to operate and maintain,
(2) Controllable with respect to pulse and stroke volume,
(3) Fabricated out of transparent materials to allow visual observation of performance. (Impending ventricle failures can be observed.),
(4) Sterilizable preferably by autoclave methods,
(5) Functional with negligible heat liberation,
(6) Fabricated out of materials compatible with the blood,
(7) Easily transportable,
(8) Sealable to preclude blood contamination and air embolism,
(9) Inexpensive to manufacture,
(10) Operable with low audible noise levels.

Existing Designs

Existing pumps and their performance capabilities offer to the designer a valuable source of experience and knowledge. Their advantages and disadvantages are considered in this section.

Present extracorporeal pumps may be classified with respect to the type of flows they produce, i.e., pulsatile or essentially nonpulsatile. Nonpulsatile pumps are the more uniform in design. They usually consist of a plastic or rubber tube or sleeve wrapped around or within a nonflexing cylinder. A roller driven by an electric motor squeezes the tube as it rolls around the cylinder. Blood is forced ahead of the roller. This type of pump has the following advantages:

1. Valves are eliminated
2. Minimal propelling pressures exist
3. Control mechanisms are relatively simple

Disadvantages may be:

1. A relatively short ventricle life
2. A lack of compactness

Pulsatile pumps are less uniform in design. They usually consist of a plastic or rubber ventricle squeezed by electric motors, pistons, electromagnets or pressurized fluids. Valves are required to control the blood flows. Such pumps have the following advantages:

1. Compactness
2. Long ventricle life
3. Pulse duplication potentialities

Disadvantages may be:

1. Excess propelling pressures
2. Valves which are both hemolytic and short lived
3. Relatively complicated control mechanisms

Exceptions certainly exist for both classes of pumps. Consideration of existing designs was found valuable in the development of the fluid amplifier pump particularly with respect to factors affecting hemolysis and ventricle life.

The Fluid Amplifier Pump

Figure 2 shows a recent R&D fluid amplifier heart pump. It consists of a fluid amplifier attached to a housing containing a plastic ventricle and two plastic tricuspid semilunar valves. Two gages attached to the housing measure the amplifier and power supply pressures. Control valves to regulate stroke volume and pulse rate have also been introduced as well as a control to adjust the flow range to the resistance of different blood circuits. These permit adjustment of the pump's output to the individual needs of the subject prior to or at the beginning of the pump run. (If the pump was intended to perfuse only a single subject with known peripheral resistances, the essential elements for satisfactory operation would be those shown in Figure 3. In this case the amplifier would be designed to accommodate the needs of the subject and the power supply without valves.)

a. Operational Description

With the aid of the pump schematic shown in Figure 4, operation may be described as follows:

A fluid (air) enters the fluid amplifier after passing through a valve (the stroke volume control-A) which regulates the amount of air flowing. Upon leaving the power nozzle as a high velocity stream, the flowing air begins to entrain additional air. Because the walls adjacent to the power jet in the interaction region - the region bounded by the power and control

jets and receivers - are not symmetrically spaced with respect to the power jet, entrainment on the side with the closer wall is impeded. Because replenishment of air is more difficult a drop in ambient pressure results. The power stream is forced to deflect slightly toward the closer wall as a result of the unbalance in pressures. Replenishment of air on this closer side is further impeded with deflection of the power stream creating still lower ambient pressures in this region and greater stream deflections. Eventually the power stream is caused to "lock on" to the closer wall. The pump will start to pulse regardless of the wall chosen by the stream to lock on initially.

If the power stream locked on to the left wall upon admission to the interaction region, it would subsequently flow through the left receiver into the housing where the increase in air pressure squeezes the ventricle, as shown in the side view in Figure 4, until the upper port is uncovered. Air then flows through the deflection control line and by virtue of an exchange of momentum between it and the power stream, the power stream is deflected to the right receiver. In this process of squeezing the ventricle by air pressure, a cardiac systole has been duplicated.

After power stream switching has occurred, blood, forced by gravity, begins to fill the ventricle and in so doing closes the deflection port. The entrainment process begins with air being drawn both from the now closed deflection control line and through the left receiver from around the ventricle. As long as the ventricle is expanding, air is supplied in sufficient quantities to satisfy entrainment, and the power stream remains locked to the right wall. When the ventricle has been completely filled, air replenishment is impeded on the left side of the power stream. In the opposite control jet the flow resistance control (B) permits a metered amount of air to enter to prevent entrainment from dropping the pressure below a predetermined level. A pressure difference builds up between the control jets. When it becomes sufficiently great, a flow occurs in the right control jet great enough to deflect the power stream back to the left receiver. The cycle is now complete.

The pulse rate control (C) in the right receiver serves to create a variable load for the amplifier, and, as a consequence, pulse rate can be controlled. The fixed bleeder bypassing the pulse rate control precludes back pressures from building up great enough to switch the power stream. Control is achieved over the pump's stroke volume by metering the amount of air entering the ventricle chamber. The stroke volume control performs this function. Adjustment of either the pulse rate or stroke volume control affects the other's function to some degree. To compensate for increased flow resistances presented to the pump, the flow resistance control, by regulating entrainment in the right control jet, allows receiver pressures to build to higher values before the power stream switches on back pressure. The flow resistance control, the fixed bleeder, and

the pulse rate control open to atmospheric pressure.

Output blood flows are made to vary directly with filling pressures by causing corresponding changes in the filled volume of the ventricle. By limiting the residual volume of the ventricle in systole, the desired effect is produced. This characteristic provides a regulation of filling pressures without electronic or other level sensing devices. Because of energy conservation, the output flows vary inversely with the flow resistances presented to the pump.

Minimal propelling pressures are realized by controlling entrainment in the right control jet and by having the power stream control both stroke volume and pulse frequency. The fact that stroke volume and pulse are controlled by the power stream offers the possibility of exploiting with a fair degree of precision the elastic responses of the vascular system by duplicating stroke volume and pulse for any particular value of flow resistance and rate. The present pump approximates this over a very narrow range of flow.

Atrial contraction during diastole can be approximated by having the power stream entrain air through the left receiver from around the ventricle to produce slightly negative pressures.

An unusual characteristic of the pump is that a pumping mode similar to ventricular fibrillation can be produced. For flow conditions of the power stream above certain energy levels (these levels can be varied), the power stream switches from receiver to receiver rapidly at about fibrillation frequencies, probably because of excessive receiver back pressures. Blood flows fall to zero.

b. Reliability and Life Characteristics

The fluid amplifier pump design should result in improved reliability and longer life for the following reasons:

(1) Apart from ventricles and valves there are no moving parts to wear or electronic components to fail. The passageways are too large to clog easily.

(2) The basic fluid amplifier components can be reproduced with consistent accuracy - possibly photographically.

(3) Maintenance is minimal. Lubrication is unnecessary.

The ventricle has been designed for negligible stress. One ventricle has been tested at a two liter per minute

water flow and 190 mm Hg blood pressure equivalent for 1,500 hours without failure.

c. Packaging

The packaging aspects of the pump are excellent. It consists of thirteen basic parts exclusive of assembly screws, pressure gages and valves. Apart from the fluid amplifier, dimensional tolerances are large. Both the housing and amplifier can be fabricated out of clear sterilizable plastics. Molding techniques can be used in the production of all pump elements. If required, the ventricle and valves can be thrown away after each use because of their low cost.

The complete pump weighs 7-3/4 pounds and has overall dimensions of 7-5/8 x 6-3/4 x 5 inches. It produces a muffled audible pulsing noise when operating. Quantity production costs for the complete pretested pump in its final design should be quite low when compared to the costs of available pumps.

Performance Results

Neither the engineering nor medical evaluation tests of the pump's capabilities have been completed. The pump would appear to have the following performance characteristics:

1. It can produce blood pressures from near 0 to about 500 mm Hg.

2. Blood flows of about 1/2 liter/minute to nine or ten liters/minute are possible depending on circuit flow resistances.

3. Pulse rate can be varied from about 20 cpm to 125 cpm. The upper limit depends on flow resistance.

The hemolysis rate appears to be less than that caused by two commercial pumps employed in comparison tests. The promising features of design discussed above have proved functional and adequate during both partial and total heart bypass proof-test experiments with animals.

It has been observed that autonomous changes in the peripheral resistance of the animal cause the output of the pump to change accordingly without adjustment of controls. Increases in peripheral resistance cause decreases in the pump's pulsing frequency and vice versa. If the increase in circulatory resistance is great enough, the pump stops pulsing altogether. It starts to pulse again when the animal's resistance has lowered to a preset level.

Decreases in filling pressure cause the pump to pulse rapidly with negligible output flows - a condition similar to

ventricular fibrillation. When the filling pressures return to a predetermined level, the pump automatically resumes normal pulsing.

Through adjustment of propelling pressures (the flow resistance control - Figure 4), the duration of the pump's systole can be varied. This reflects in the maximum systolic-minimum diastolic blood pressures and these changes have been observed in the arterial system of an animal.

Problem Areas

Apart from design improvements intended for future production models, two design problems need better solution in the R&D prototype described herein. One problem concerns valves. The life of the tricuspid semilunar types produced for the pump is relatively short, although their hemolytic characteristics are good. Other valve designs are being considered. The second problem involves an insufficient range of frequency adjustment when high resistance catheters are used with the pump. For arterial blood pressures of less than 200 mm Hg the upper frequency limit is about 100 cpm. This limit drops below 60 cpm for catheters less than 16 French in diameter. If resonance phenomena are important in the cardiovascular system, this upper frequency limit needs to be raised. This can be done by reducing stroke volume.

No attempt has been directed toward causing the pump of Figure 2 to synchronize its pulses with those of the subject's heart. This may pose future problems.

Conclusion

A heart pump using fluid amplification principles has been designed that can approximate many of the human heart's pumping functions. It operates without moving control parts or electronics. Without such parts, reliability and life are measurably increased and production costs significantly decreased. Pumping and hemolytic characteristics are at least equal to those of available blood pumps.

Of particular importance to the Army in the field is that the pump can be transported by hand, it can be completely disassembled and reassembled in minutes, and except for ventricles and heart valves all parts can be autoclaved. Its rugged construction should permit good toleration of rough handling. Power can be provided by tanks of compressed gas or eventually perhaps even by exhaust gases from internal combustion engines.

List of References

1. Army Research and Development Newsmagazine, Office of the Chief, Research and Development, Headquarters of the Army, Washington 25, D. C., Vol. 2 No. 2 February 1961.

2. Woodward, Kenneth E., Barila, Timothy G., and Nunn, Daniel E. "Some Engineering Considerations of the Human Cardiovascular System", DOFL TM-62-5 dtd 7 February 1962.

3. Best, Herbert Charles and Taylor, Norman Burke, "The Physiological Basis of Medical Practice", The Williams and Wilkins Company, Baltimore, Md., 1961.

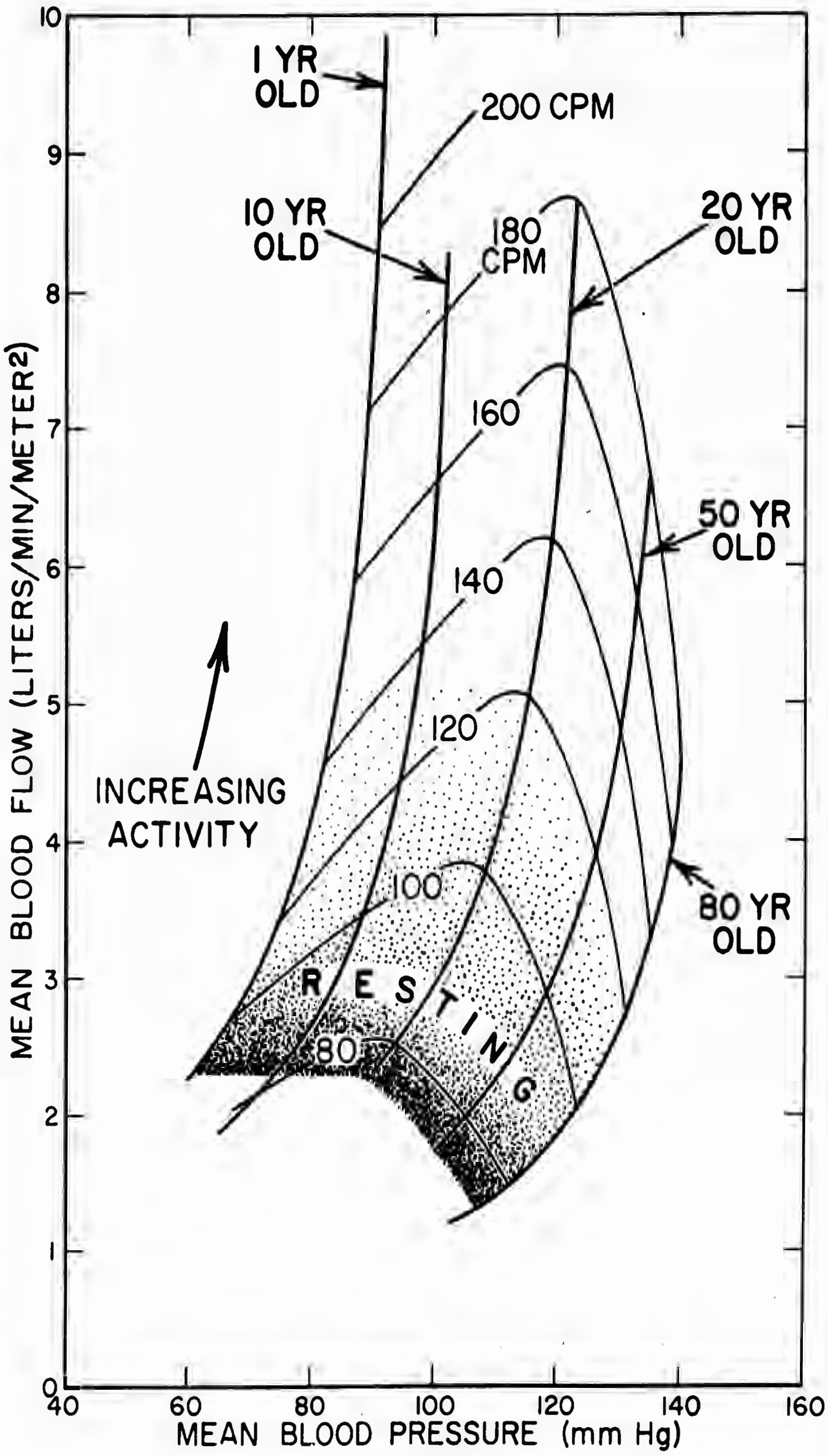


FIGURE 1.

CARDIOVASCULAR CHARACTERISTICS

(THE HUMAN MALE)

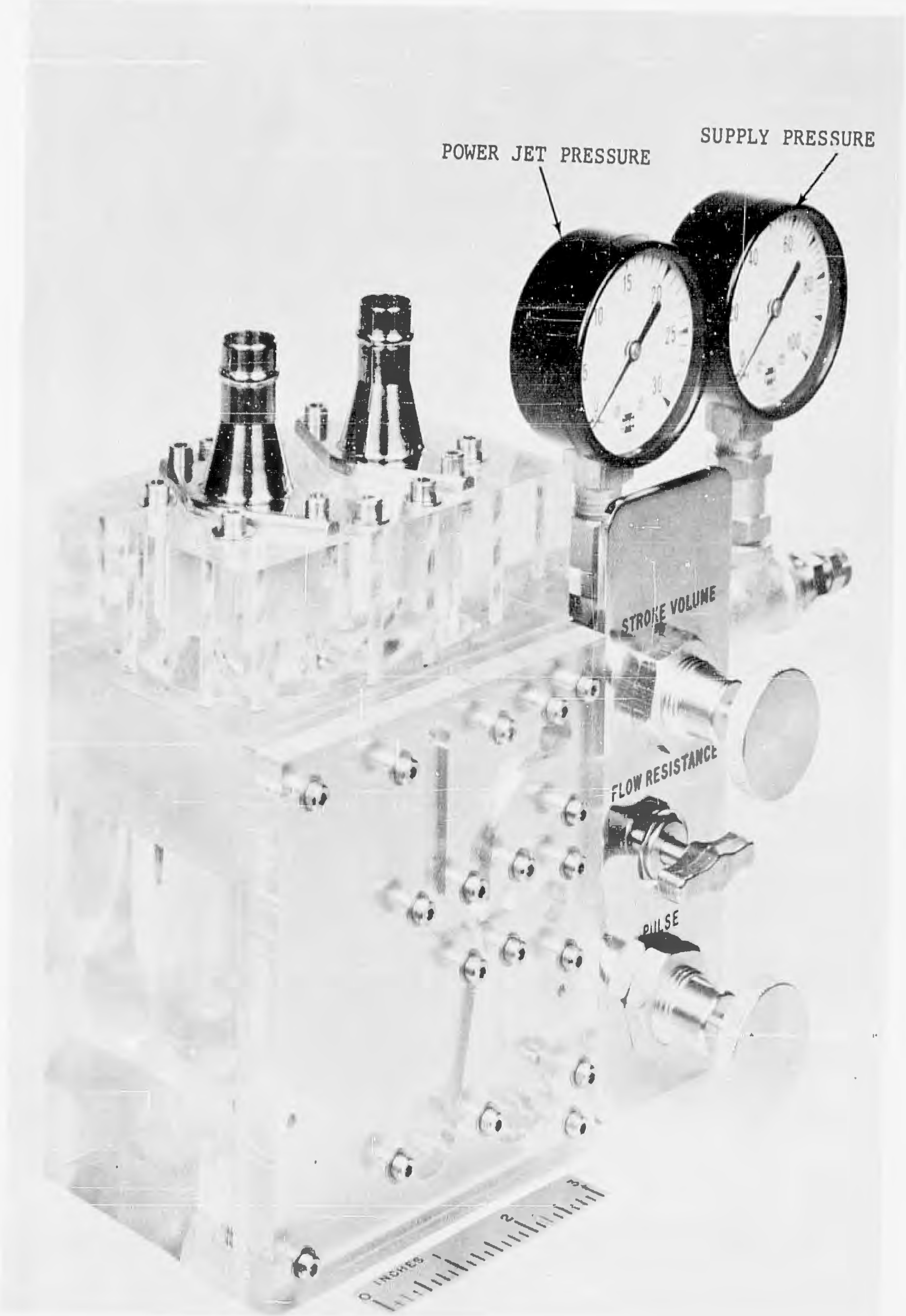


FIGURE 2
THE FLUID AMPLIFIER HEART PUMP

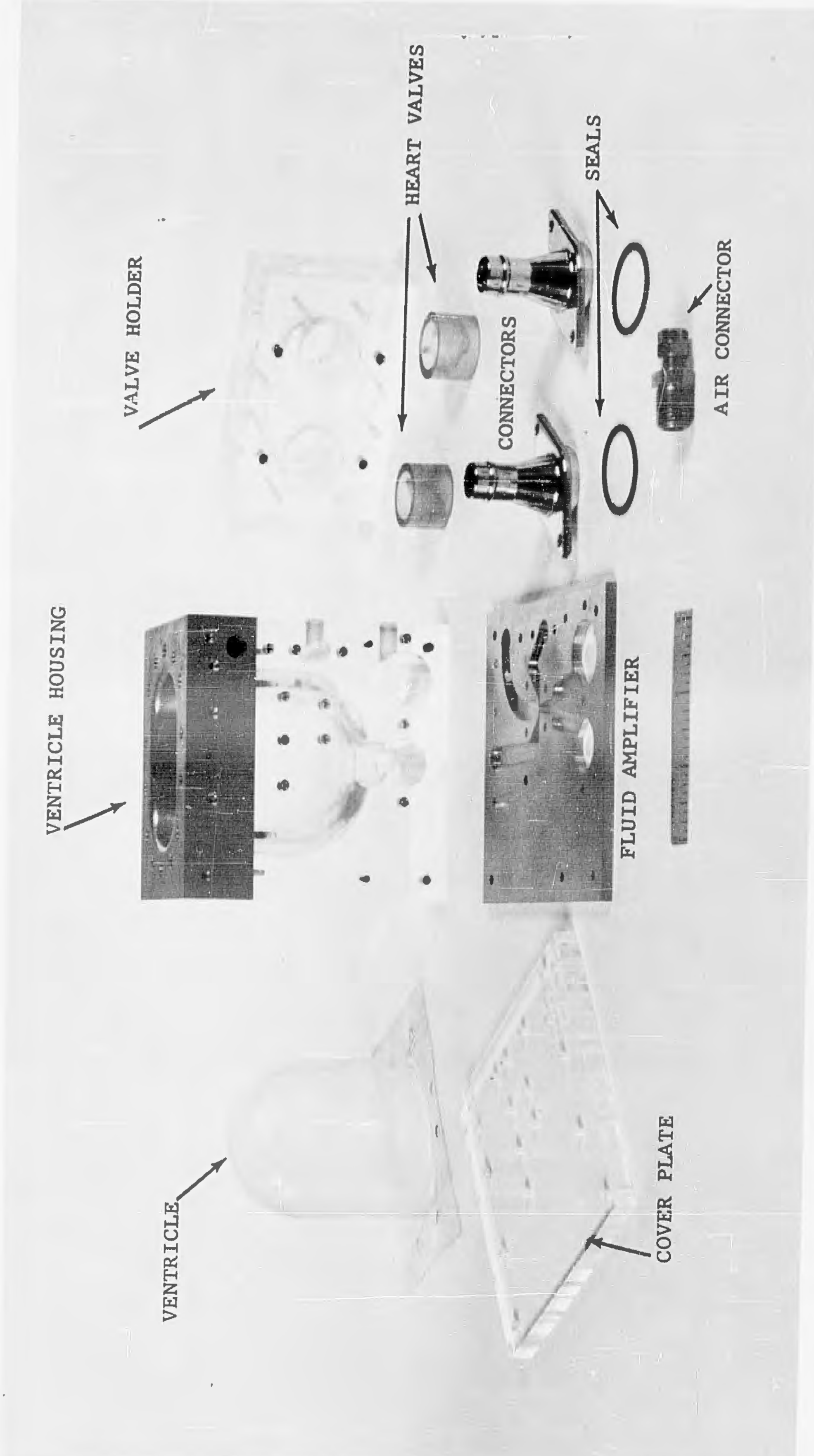


FIGURE 3
BASIC PUMP ELEMENTS

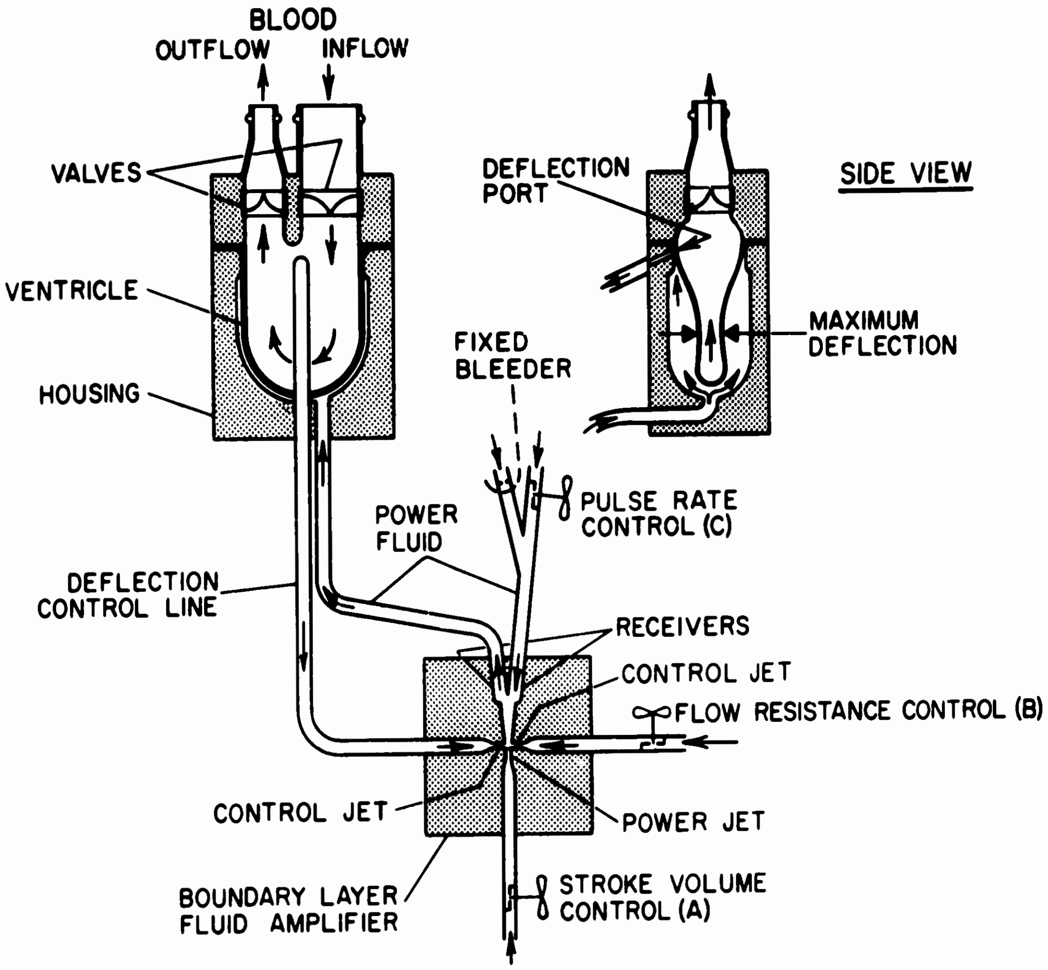


FIGURE 4.

HEART PUMP SCHEMATIC

EFFECT OF FOOD INTAKE ON PERFORMANCE CAPABILITIES

DONALD R. YOUNG
QM FOOD AND CONTAINER INSTITUTE FOR THE ARMED FORCES
QM RESEARCH AND ENGINEERING COMMAND, U.S. ARMY
CHICAGO, ILLINOIS

A belief that an unusual state of robust health and increased work performance capacity might be attained by special nutrition has stimulated man in his quest for salubrious diets (1) since antiquity. The idea that performance is somehow commensurate with the food we consume is certainly not new. It was implied, in fact, in an aphorism of Hippocrates: "In which cases is food to be given once or twice a day, and greater or smaller quantities; something must be conceded to habit, season, country, and age." It would appear that Hippocrates had pondered, at least, the significance of amount and time of food intake on physical well-being.

The problem of the relationship of nutrition to performance is not easily answered. Consistent improvement of performance capacity by abundant supplementation with various known nutrients (1, 2, 3) has not been demonstrated. Some reports indicate, in fact, that performance is stable even during an austere period of undernutrition that leads to an 8 - 12% body weight loss (4, 5).

Much of the available information regarding nutrition and performance is contradictory. Typical of the conflicting reports are the studies with phosphates, quoted by Keys (1): Apparently confirmatory studies conducted with police officers in 1919 led to the widespread use of foods and drinks rich in phosphates to improve muscular endurance and for a large variety of disabilities. Despite a growing skepticism, scientific tests were continued until 1942. At present, the underlying assumption that intake of phosphorus leads to improved physical capacity is subject to considerable question. But in the meantime it is interesting to note that some college athletes are rediscovering the early reports and strongly advocating phosphates to allay fatigue and increase endurance. Similar conflicting reports (3) may be found on the subject of special vitamin, protein and amino acid supplements.

A study of performance in relationship to nutrition is, however, much more than a search for single nutrients which individually or when fed in combination might enhance human capabilities. There are other important parameters. In our approach (6) we have included studies of such factors as effect of environment, frequency and periodicity of feeding (7), food deprivation (8), over-feeding (9), and effect of water and special nutrient supplements (10, 11); we have combined these studies with basic studies of physiology and metabolism during the stress of physical work (12, 13, 14, 15, 16). In addition, since we feel that mental alertness and judgment ability are important components of fitness, we have investigated these variables as well. The capacity for gross physical work does not alone, we believe, give the complete picture.

Although our interest is principally in human performance, the exploratory nature of the studies and the stress and strain of the tests has required that many of the experiments be undertaken with experimental animals. Perhaps the most important finding has been that recent intake of food can actually impair physical work capability. This finding, of course, requires some elaboration. Specifically, our evidence supports the conclusions that (a) where physical exertion is involved, the body responds best after a period of fasting, but (b) where mental processes are concerned, there is a decrease in efficiency as fasting progresses.

EXPERIMENTAL

All tests described below were conducted with male, pure-bred beagle dogs, 1 - 2 years of age, and weighing 8 - 12 kg.

Studies of physical work capacity were undertaken in the Performance Laboratory, Quartermaster Food and Container Institute for the Armed Forces, Chicago, Illinois. These tests followed a common plan of aerobic work on treadmills driven at a speed of 3.6 mph, figures 1 and 2. The degree of incline was adjusted for individual differences in body weight in order to establish a constant work load of 200 kg-m/min. Typically, the environment provided a dry bulb temperature of 65° F., 50% R.H., with a wind velocity (predominantly laminar flow) of 100 ft/min. Lighting and noise level were held constant. Work capacity expressed in kilo-calories was determined as the product of total running time and rate of energy expenditure. The dogs were considered to be exhausted when they refused to continue running after 6 consecutive shocks from an electric barrier mounted at the rear of each treadmill.

During work, respiratory gas exchange was determined by using specially constructed respiratory masks for collection of a series of 2-minute samples of expired air in chain compensated gasometers. Rubber respiratory face masks were fitted to each dog

and designed to permit minimum discomfort and maximum ventilation during work. The mask covered the snout and extended well past the angle of the jaw. An internal face flange padded with foam rubber sealed effectively the commissures of the lips within the mask without interfering with vision, swallowing, or circulation to the lips and soft tissues of the muzzle. The internal valve system minimized the respiratory dead space. O_2 consumption and CO_2 production were determined by analysis of expired air samples in the Beckman Oxygen Analyzer and Haldane apparatus, respectively. Calculations of energy expenditure were made from the tables of Zuntz (17).

Blood and urine samples collected during the periods of observation were stored and analyzed for various metabolites by standard procedures (18). Usually six to eight dogs were used for every experiment, and each animal served as his own control. This provided sufficient data to test statistically for significant differences and treatment effects.

Studies of sensory discriminative ability were conducted under contract (DA 19-129-QM-1539) by Dr. Jack Arbit, Northwestern University, Chicago, Illinois. The tests were conducted within a small compartment constructed in a soundproof environmental chamber. All studies were done at ambient air temperatures (70° - 76° F.). The visual stimuli consisted of 7 white six-watt bulbs arranged horizontally. The bulbs were wired in a manner which permitted their presentation in any combination. In addition, they were wired to a rheostat so that light intensity could be controlled. Simultaneous with the onset of the lights, auditory cues were presented and a timer was started. When the animal depressed a bar, the sound and lights were extinguished automatically, the timer was stopped, and food (small bits of mackerel) was dispensed to the animal. In most of the tests, the center light with a high tone (2500 cps) was arbitrarily designated the rewarding stimulus. Responses to the peripheral lights with a low tone (1500 cps) were unrewarded or resulted in electric shock. The light signals were programmed to appear in a random order and the duration of each light stimulus was 10 seconds. Four to six animals were used in each experiment, and each animal served as his own control.

RESULTS

Physical Work Capacity

Physiologic responses to work. A comparison of the responses of sedentary dogs, restricted to their respective cages for 20 hours, and active dogs running continuously for 12 hours is shown in table 1. In working dogs, the average gross energy expenditure was 2100 Cal and approximately 3.5 times greater than that measured in sedentary animals. Increased heart rate and respiratory rate

reflect the differences in the states of physical activity. It will be noted that there is no oxygen debt nor is there an accumulation of blood lactic acid in the work test we employ. Physical activity modified only three of the relatively large number of components examined in blood and urine samples. Thus, continuous running depressed the blood glucose, increased the blood level of non-protein nitrogen, and increased urinary potassium excretion. The differences are significant at the 1% level of confidence (p = .01). The results indicate (a) a tendency towards a depletion of the carbohydrate reserves during work, (b) an increased breakdown of body tissue leading to increased blood levels of nitrogen; the nitrogen, however, is retained by the body and not lost in the urine, and (c) urinary elimination of the potassium liberated during tissue breakdown.

Metabolism and water requirements during work. Basic studies were undertaken to determine the metabolic requirements during work. Energy metabolism was examined during sub-maximal work following the intake of a standard meal (Purina Chow) known to maintain daily body weight balance. It will be noted in table 2 that during work administered 4 hours following the intake of 600 kilocalories, the percentage of calories derived from carbohydrate, fat, and protein oxidation is 70, 24, and 6%, respectively. Seventeen hours after feeding, the percentages of energy derived from carbohydrate, fat, and protein are significantly different, $p < .01$, and in the ratio of 48, 50, and 2%, respectively. Thus, in the postabsorptive state, the calories derived from the oxidation of carbohydrate, fat, and protein are in the ratio 48/50/2; feeding a meal relatively high in carbohydrate and protein increases carbohydrate and protein oxidation and reduces fat oxidation.

Since dehydration is a principal cause of fatigue, studies were initiated to ascertain the water requirements and effect of variations in fluid intake on the performance test scores. For this purpose, the animals were allowed to drink from stainless steel pans during exhaustive work. The effect on work capacity of consuming 0, 480, 820, 1140, and 1675 cc of water during treadmill running was studied. These data are shown on a semilogarithmic plot in figure 3. The relationship between water intake and work capability was predominantly linear; the product-moment coefficient of correlation between the two variables is +.85. Additional tests showed that provision of water in excess of 2 l. did not increase further running capability. On the basis of the results presented, it is concluded that (a) provision of approximately 1.2 - 1.6 l. of water is required to exploit the full work potential of the animal, and (b) during work, factors other than available drinking water will limit performance eventually.

Relationship between work capacity and food intake. Studies were undertaken to follow maximum work capacity during treadmill running administered at various periods following the consumption of

a standard diet (Purina Chow). The earlier metabolism studies had shown that intake of food provided the necessary fuel for muscular work and spared the body stores of pre-formed carbohydrate. Furthermore, roentgenographic studies showed that at 4 to 6 hours after feeding, most of the ingested food had passed into the large intestine, and therefore it was likely that digestion and absorption had been completed. On the basis of our knowledge of water requirements, we felt that at 6 hours after feeding and with an unlimited supply of drinking water available, the animals would be in the most advantageous physiologic state for maximum work capacity.

Separate running tests were administered 2, 4, 6, 17, and 120 hours after intake of a 200 g. meal. The relationship between work capacity and recency of feeding is shown in table 3. During the first 6 hours after feeding, the average work capacity was 750 Cal with a standard deviation of 275. Seventeen hours after the last meal, work capacity was 2200 Cal and significantly ($p < .01$) higher than values measured during the preceding periods. Finally, at 120 hours (5 days) after the last meal, work capacity was 4100 Cal and significantly ($p < .01$) greater than all other scores. Indeed, during the latter trials one dog ran 140.4 miles uphill over a period of 38.7 hours. We feel this is an unofficial record for continuously sustained performance.

To establish further a relationship between work performance and the foods consumed, additional tests were undertaken to identify specific nutrients associated with decrement in performance capacity. In these studies, whole milk with added copper and iron was used as the control to study the comparative effect on performance of providing either no supplement, water, or separately pure solutions of protein, carbohydrate, fat, cellulose derivatives, phospholipids, or water soluble vitamins to postabsorptive animals. These substances are common to all diets composed of natural foods; they were given in aqueous solutions to prevent the dehydration usually occurring during work. With the exception of the solution containing 1.5% carboxymethylcellulose, all other fluids were similar to milk in their content of: a) protein; b) carbohydrate; c) fat; d) phospholipids; and e) niacin, pantothenic acid, riboflavin, thiamine, and pyridoxine. The nutrient composition of the solutions tested is shown in table 4. Supplementation with cold solutions was initiated after 1 hour of running and continued on an <u>ad libitum</u> basis until the termination of work.

The percentage change in performance as a result of supplementation with the various nutrients is shown in table 5. Total fluid intake during the individual trials was similar (1 l.). Average work capacity of animals consuming milk, and solutions of corn oil, vitamins, phospholipids, and carboxymethylcellulose was 901 $\pm$ 287 Cal. In contrast, animals consuming only water, or solutions of protein or glucose showed a maximum work capacity of 1700 $\pm$ 326 Cal. The

difference in the means is significant ($p = .01$). The results indicate that consumption of whole milk, fat, vitamins, phospholipids, and cellulose derivatives during periods of physical activity can impair performance capacity. Glucose and protein are apparently neither beneficial nor detrimental.

Extensive blood studies were undertaken to determine the mechanisms whereby intake of the various nutrients might affect performance. In general, however, there was no systematic relation between the blood components (glucose, fructose, pyruvate, urea) studied and work capacity. Surprisingly, the level of blood sugar was unaffected by intake of glucose. For example, in dogs consuming only water during work, the final level of blood glucose was 51 ± 5 mg.%; in dogs fed the sugar solution, the level of blood glucose was 57 ± 11 mg.%. With the exception of a lactosemia and galactosemia found in the milk-supplemented animals, most of the blood components were in the normal range as shown in table 1.

Similarly, there were no significant differences in the respiratory or cardiovascular variables.

Rapid rise in body temperature, however, appeared to be associated with decreased work capability. For example, for all tests, running time varied from 175 to 920 min.; rectal temperature showed a systematic, although nonlinear, rise during treadmill running and at the termination of work was 41° C. or approximately 2° C. higher than the resting temperature. Thus, the animals showing poor performances (the shorter runners) attained elevated body temperatures more rapidly than did the longer runners. Preliminary studies employing dogs with fistulae and thoracic duct cannulae suggest that intake of certain solutions, particularly fat, results in decreased intestinal motility and water absorption. Therefore, it is possible that the effects we have observed are due to failure of water absorption and a resulting functional dehydration. Additional studies are in progress to test this hypothesis.

Psychomotor Performance

Effect of food deprivation on sensory discrimination. Systematic studies were undertaken to follow stimulus generalization over various periods of food deprivation. The results of earlier tests indicated that during the stress of induced anxiety and fear or 72 hours of food withdrawal, discriminatory responses were significantly modified in that the animals responded more frequently to unrewarding (incorrect) stimuli. In consequence, we felt that additional tests were required to evaluate more fully the breakdown in discrimination as well as the perceptual-motor processes.

Typically, the animals were maintained on a standard laboratory dog ration and examined at various periods following the

last meal. The effect of 1, 2, 4, 8, 17, 72, and 120 hours of food deprivation on frequency of responses and reaction time is shown in table 6. Here we see a relative constancy in responses and reaction times during the first 17 hours of food withdrawal. Thus, during this period the average frequency of incorrect responses and reaction time was 19% and 7.2 sec., respectively, with very little variation from the mean. At 72 hours of deprivation, the dogs responded more frequently (35%) and faster (3.2 sec.) to incorrect cues. Finally, after 5 days without food there was a 100% response to the unrewarding stimuli and reaction time was less than 0.1 sec. The latter score differs significantly ($p < .01$) from the results obtained during the first day of food withdrawal. From these data it may be concluded that (a) hunger of 1 to 17 hours duration has no effect on judgment ability, (b) disorganization of behavior begins after approximately 3 days of food deprivation, and (c) is particularly marked after 5 days without food.

Further tests were undertaken to examine intensity discrimination under normal conditions and during 72 hours of food deprivation. The response times (for rewarding stimuli) to five different light intensities ranging from the brightest to full off are shown in figure 4. It will be noted that there was a general tendency for the stressed animals to respond more quickly to rewarding stimuli of graded intensity. Thus, unlike the finding of breakdown and disorganization in selective behavior for spatially oriented cues, intensity discrimination showed a relative sharpening under stress conditions.

CONCLUSIONS

The studies reported above agree in demonstrating an important relationship between food intake and performance capability. Two problems are recognized with regard to gross physical work. Physical capacity is reduced by recent intake of a nutritionally balanced meal and by selected components of natural foods. Factors tentatively associated with impaired capabilities are elevated body temperature and systemic dehydration as a result of delayed water absorption. Secondly, a more basic problem of tissue adaptation to food deprivation is implicated. Thus, 17 hours after feeding when the intestinal tract is empty, work capacity is 2200 Cal; after 5 days of food deprivation work capacity is increased to 4100 Cal. The permissive effect of such treatment is of considerable interest, and the search for the underlying mechanisms of action is underway.

Judgment ability is modified in a different direction by food withdrawal. As the period of fasting progresses, the responses become more rapid and the probability of error increases. Our tests show a general breakdown in discrimination ability after 3 days without food which becomes increasingly severe on the 5th day.

Ability to modify performance by control of food intake and dietary regimen has obvious appeal. For example, feeding schedules may be developed to maximize performance for particular tasks, mental or physical. However, before the full potential of maximizing performance by dietary control can be realized, more research is necessary to establish the validity of our initial observations for man.

Certain problems that arise in feeding active subjects have been presented. These problems must be answered if the advantages assured by a high level of performance capacity are to be attained. Not to seek the answers is to assume that the limitations imposed by diet on performance can be overcome by sheer will power. Such an assumption is invalid since a fit subject can be identified from among his companions and his operating characteristics appraised in the most rigorous terms.

LITERATURE CITATIONS

1. Keys, A. Physical performance in relation to diet. Fed. Proc., 2: 164, 1943.

2. Simonson, E. Influence of nutrition on work performance. Nutrition Fronts in Public Health No. 3, p. 72, 1951.

3. Young, D. R. and H. Spector. Physical performance capacity and nutriture: Evaluation of rations by animal experimentation. Am. J. Clinical Nutrition, 5 (2): 129, 1957.

4. Henschel, A., H. L. Taylor, and A. Keys. Performance during acute starvation. J. Appl. Physiol. 6: 624, 1954.

5. Keys, A., J. Brozek, A. Henschel, O. Mickelsen, and H. L. Taylor. The Biology of Human Starvation, Vol. 2. The Univ. of Minnesota Press, Minneapolis, Minn., 1950.

6. Young, D. R. Performance in relationship to food energy intake. Activities Report, Res. and Dev. Associates, Food and Container Institute, Inc., 13: 146, 1961.

7. Young, D. R., A. Iacovino, P. Erve, R. Mosher, and H. Spector. Effect of time after feeding and carbohydrate or water supplement on work in dogs. J. Appl. Physiol., 14 (6): 1013, 1959.

8. Young, D. R. Effect of food deprivation on treadmill running in dogs. J. Appl. Physiol., 14 (6): 1018, 1959.

9. Young, D. R. Effect of body composition and weight gain on performance in the adult dog. J. Appl. Physiol., 15 (3): 493, 1960.

10. Young, D. R., N. S. Schafer, and R. Price. Effect of nutrient supplements during work on performance capacity in dogs. J. Appl. Physiol., 15 (6): 1022, 1960.

11. Young, D. R. and R. Price. Effect of lipid supplements during work on performance capacity in dogs. Interim Report, QMF&CIAF, Rpt. No. 15-61, July 1961.

12. Young, D. R., R. Mosher, P. Erve, and H. Spector. Body temperature and heat exchange during treadmill running in dogs. J. Appl. Physiol., 14 (5): 839, 1959.

13. Young, D. R., R. Mosher, P. Erve, and H. Spector. Energy metabolism and gas exchange during treadmill running in dogs. J. Appl. Physiol., 14 (5): 834, 1959.

14. Young, D. R. A comparison of energy expenditure in man and dog at rest and during work in the normal and fasting state. Interim Report, QMF&CIAF, Rpt. No. 7-60, Feb. 1960.

15. Young, D. R. and R. Price. Utilization of body energy reserves during work in dogs. J. Appl. Physiol., 16 (2): 351, 1961.

16. Young, D. R., R. Price, N. R. Elder, and R. R. Adachi. Energy metabolism and urinary nitrogen, electrolytes, norepinephrine, and corticosteroids in relationship to work in dogs. J. Appl. Physiol. In press.

17. Best, C. H. and N. B. Taylor. In: The Physiological Basis of Medical Practice (2nd Ed.) Baltimore: Williams & Wilkins, 1939, p. 852.

18. Biochemical and dietary procedures. U. S. Army Medical Research and Nutrition Laboratory. Rpt. No. 242. Fitzsimons General Hospital, Denver, Colorado, 29 Jan. 1960.

Table 1. Physiologic Characteristics of Work, Blood Chemistry and Urinary Excretion in Postabsorptive Dogs

	Sedentary, 20 hrs	Active, 12 hrs
Energy Expenditure, Cal/min	0.50 ± .04	3.35* ± 0.3
Gross Energy Expenditure, Cal	600 ± 30	2100* ± 210
Heart Rate/min	90 ± 15	235* ± 12
Respiratory Rate/min	25 ± 4.0	210* ± 15
Venous O_2 cc/100 ml	16 ± 1.0	15.9 ± 0.8
Oxygen Debt	None	None
Blood Glycogen, mg%	2.2 ± 1.5	2.8 ± 1.6
Blood Glucose, mg%	72 ± 5.0	55* ± 6.0
Blood Fructose, mg%	3.3 ± 0.9	2.8 ± 0.6
Blood Pyruvate, mg%	2.1 ± 0.5	2.0 ± 0.4
Blood Lactate, mg%	3.0 ± 1.5	4.0 ± 0.8
Blood NPN, mg%	31 ± 5.2	50* ± 10.0
Blood Acetone, mg%	2.0 ± 0.7	3.0 ± 1.4
Urinary NPN, gm	2.2 ± 0.7	1.89 ± 0.6
Urinary Urea N, gm	1.69 ± 0.4	1.50 ± 0.5
Urinary Creatinine N, mg	75 ± 32	57 ± 13
Urinary Acetone, mg	12 ± 5.0	18 ± 7.8
Urinary Sodium, Meq	22 ± 3.0	24 ± 3.0
Urinary Potassium, Meq	12 ± 2.5	18* ± 2.5
Urinary Phosphorus, mg	195 ± 44	183 ± 30

Values are means ± Standard Deviations.
* Differ significantly from values obtained with sedentary dogs, $p \leq .01$.

Table 2. Energy Metabolism During Work in Dogs Following Intake of a Standard Mixed Diet[1]

	W4* (600 Cal)	W17* (600 Cal)
Respiratory Quotient	.91 ± .07	.85 ± .05
Calories from CHO, %	70 ± 10	48** ± 15
Calories from Fat, %	24 ± 10	50** ± 16
Calories from Protein, %	6.5 ± 2.1	2.3** ± 0.6

Values are means ± Standard Deviations.
[1] Purina Chow (600 Cal), Calorie composition: Carbohydrate, 53%; Fat, 19%; Protein, 29%.
* Work at 4 or 17 hours following food intake.
** Differs significantly from values measured 4 hours after feeding, $p < .01$.

Table 3. Relationship Between Work Capacity and Recency of Food Intake in Dogs[1]

Time After Last Feeding, hrs	Work Capacity, Cal*
2	850 ± 250
4	600 ± 275
6	700 ± 300
17	2200 ± 219
120	4100 ± 300

1 200 g. Purina Chow.
* Values are means ± Standard Deviations.

Table 4. Nutrient Composition of Solutions Tested

Lactalbumin, gm/100 ml	3.5
Glucose, gm/100 ml	5.0
Corn Oil, gm/100 ml	5.0
Phospholipids, gm/100 ml	0.35
Water-soluble Vitamins, mg/l.*	
Niacin	3.0
Pantothenic Acid	3.0
Riboflavin	0.75
Thiamine	0.36
Pyridoxine	0.36

* Containing additionally Inositol 17 mg/l., Choline 17 mg/l., p-aminobenzoic acid 8 mg/l., Folic acid 0.08 mg/l., Biotin 0.02 mg/l., and B_{12} 0.02 mg/l.

Table 5. Effect of Nutrient Intake During Treadmill Running on the Work Capacity of Dogs

Nutrient	Percent change in work capacity*
Milk	-61
Corn Oil	-38
Water-soluble Vitamins	-37
Plant Phospholipids	-40
Carboxymethylcellulose	-60
Glucose	<1
Lactalbumin	+ 4

* Compared with consuming only water.

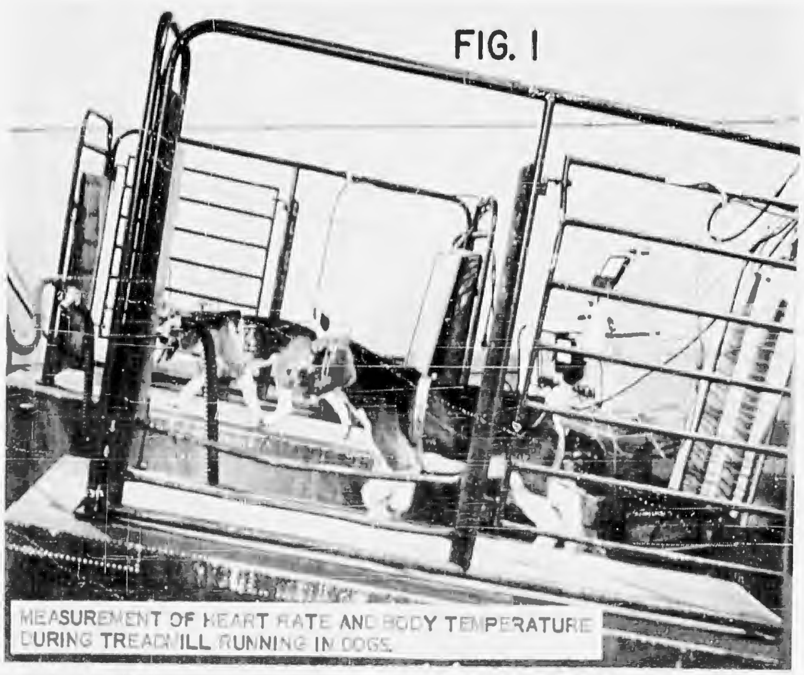


FIG. 1

MEASUREMENT OF HEART RATE AND BODY TEMPERATURE DURING TREADMILL RUNNING IN DOGS.

FIG. 2

MEASUREMENT OF RESPIRATORY GAS EXCHANGE DURING TREADMILL RUNNING IN DOGS.

FIG. 3 RELATIONSHIP BETWEEN WATER INTAKE DURING WORK AND PERFORMANCE CAPACITY IN POSTABSORPTIVE DOGS

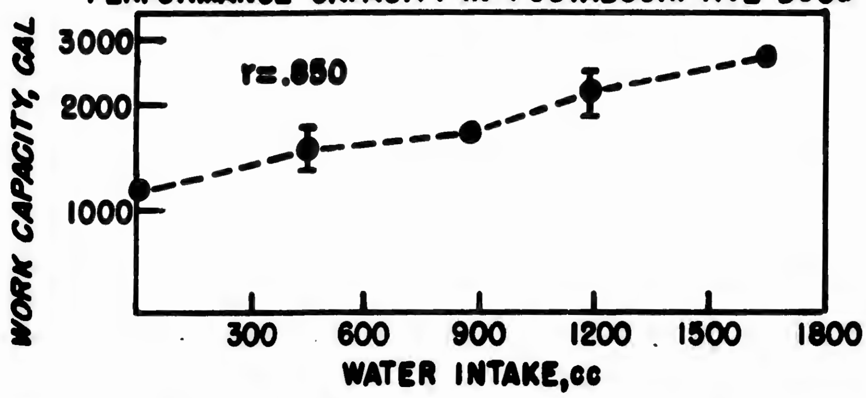


FIG.4 RELATIONSHIP BETWEEN REACTION TIME AND LIGHT INTENSITY DURING STRESS CONDITIONS

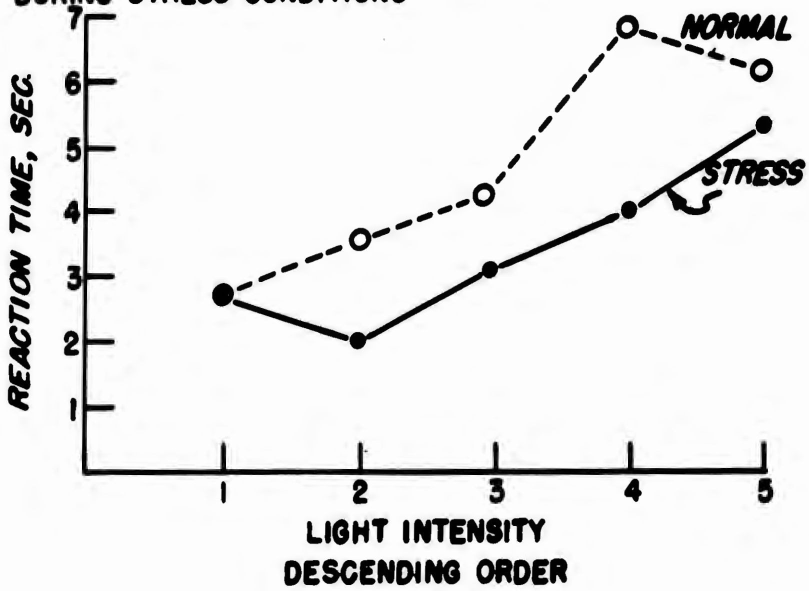


Table 6. Percent of Incorrect Responses and Reaction Times During Food Deprivation

Time after last feeding, hrs	Responses to unrewarding stimuli, %	Reaction time, Sec.
1	17	7.2
2	22	7.8
4	21	6.9
8	18	7.3
17	16	6.9
72	35	3.2
120	100	<0.1

DISTRIBUTION LIST

DEPARTMENT OF DEFENSE

Director of Defense, Research & Engineering Washington 25, D.C.	1
Armed Services Technical Information Agency Arlington Hall Station Arlington 12, Virginia	10
Chief Defense Atomic Support Agency Washington 25, D.C.	1
Director Weapons Systems Evaluation Group Department of Defense Washington 25, D.C.	1
Director Armed Forces Institute of Pathology Washington 25, D.C.	1
Commandant Industrial College of the Armed Forces Fort Lesley J. McNair Washington 25, D.C. Attn: Library	1
Commandant National War College Fort Lesley J. McNair Washington 25, D.C. Attn: Library	1
Commandant Armed Forces Staff College Norfolk, Virginia Attn: Library	1

DEPARTMENT OF THE ARMY

Assistant Secretary of the Army (R&D) Department of the Army Washington 25, D.C.	1
Chief of Staff United States Army Washington 25, D.C.	1
Chief of Research and Development United States Army Washington 25, D.C. Attn: Research Support Division	180

DEPARTMENT OF THE ARMY (Continued)

Chief US Army R&D Liaison Group (9983) (Europe) APO 757 New York, New York	1
Chief US Army Security Agency Arlington Hall Station Arlington 12, Virginia	1
Deputy Chief of Staff for Logistics United States Army Washington 25, D.C.	1
Deputy Chief of Staff for Military Operations United States Army Washington 25, D.C.	1
Deputy Chief of Staff for Personnel United States Army Washington 25, D.C.	1
Assistant Chief of Staff for Intelligence United States Army Washington 25, D.C.	1
Commanding Officer US Army R&D Group (Far East) (9984) APO 343 San Francisco, California	1
Commanding Officer US Army Personnel Research Office Washington 25, D.C.	1
Commanding Officer Army Research Office (Durham) Box CM, Duke Station Durham, North Carolina	1
Commanding General US Army Chemical Corps Materiel Command Army Chemical Center, Maryland	1
Commanding Officer US Army Biological Laboratories Fort Detrick, Maryland	1

Chemical Corps (Continued)

Commanding Officer US Army Chemical Corps Engineering Command Army Chemical Center, Maryland	1
Commanding Officer US Army Chemical Warfare Laboratories Army Chemical Center, Maryland	1
Commanding Officer US Army Chemical Nuclear Defense Laboratory Army Chemical Center, Maryland	1
Commanding Officer US Army Chemical Corps Training Command Fort McClellan, Alabama	1
Commanding Officer US Army Chemical Corps Proving Ground Dugway Proving Ground Dugway, Utah	1
Commanding Officer Rocky Mountain Arsenal Denver 2, Colorado	1
Commanding Officer Pine Bluff Arsenal Pine Bluff, Arkansas	1
Commanding Officer US Army Chemical Corps Intelligence Agency Arlington Hall Station Arlington 12, Virginia	1

Corps of Engineers

Chief of Engineers United States Army Washington 25, D.C.	1
Commanding General US Army Engineer Division Ohio River Division Laboratories Cincinnati 27, Ohio	1
Director Cold Regions R&E Laboratory Corps of Engineers Hanover, N.H.	1
Director US Army Waterways Experiment Station P.O. Box 631 Vicksburg, Mississippi	1
Director Engineer R&D Laboratories Fort Belvoir, Virginia Attn: Technical Documents Center	1

Corps of Engineers (Continued)

Director US Army Engineer Geodesy Intelligence and Mapping R&D Agency Fort Belvoir, Virginia	1
Commanding Officer US Army Engineer (USAE) Experiment Station Jackson Installation P.O. Drawer 2131 Jackson, Mississippi	1
Commanding Officer US Army CRREL Keweenaw Field Station Corps of Engineers Houghton, Michigan	1

Army Medical Service

The Surgeon General United States Army Washington 25, D.C.	1
Commanding General US Army Medical R&D Command Washington 25, D.C.	1
Commanding Officer US Army Research Institute of Environmental Medicine Natick, Massachusetts	1
Commanding Officer US Army Medical Research Laboratory Fort Knox, Kentucky	1
Commanding Officer US Army Medical Research & Nutrition Laboratory Fitzsimons Army Hospital Denver 30, Colorado	1
Commanding Officer US Army Surgical Research Unit Brooke Army Medical Center Fort Sam Houston, Texas	1
Commanding Officer Medical Equipment Development Laboratory Fort Totten, New York	1
Director Walter Reed Army Institute of Research Walter Reed Army Medical Center Washington 12, D.C.	1

Commanding General 1
US Army Ballistic Missile Agency
Redstone Arsenal, Alabama

Commanding General 1
Aberdeen Proving Ground
Aberdeen, Maryland

Commanding General 1
Aberdeen Proving Ground
Aberdeen Maryland
Attn: Human Engineering Laboratory

Commanding General 1
Army Ordnance Missile Command
Redstone Arsenal, Alabama

Commanding General 1
US Army Rocket & Guided Missile Agency
Redstone Arsenal, Alabama

Commanding General 1
Army Ordnance Missile Command
Redstone Arsenal, Alabama
Attn: ORDXM-XGR

Commanding Officer 1
Diamond Ordnance Fuze Laboratory
Connecticut Ave., & Van Ness Sts., N.W.
Washington 25, D.C.

Commanding General 1
Frankfort Arsenal
Bridge & Tacony Streets
Philadelphia 37, Penna.

Commanding General 1
Ordnance Materials Research Office
Watertown Arsenal
Watertown 72, Mass.

Commanding General 1
Ordnance Tank-Automotive Command
Detroit Arsenal. Center Line, Michigan

Commanding General 1
White Sands Proving Ground
New Mexico
Attn: Technical Library

Commanding Officer Picatinny Arsenal Dover, New Jersey	1
Commanding Officer Rock Island Arsenal Rock Island, Illinois	1
Commanding Officer Springfield Armory Springfield, Mass.	1
Commanding Officer Watertown Arsenal Watertown 72, Mass.	1
Commanding Officer Watervliet Arsenal Watervliet, New York	1
Commanding Officer Headquarters, Ordnance Weapons Command Rock Island Illinois	1
Commanding Officer US Army Ordnance Special Weapons Ammunition Command Dover, New Jersey	1
Institute for Research, 1010 Frankfort Arsenal Philadelphia 37, Penna.	1
Commanding General Quartermaster R&E Command Natick, Mass.	1
Commanding Officer Quartermaster Food & Container Institute for the Armed Forces Chicago, Illinois	1
Commanding Officer Quartermaster Field Evaluation Agency Fort Lee, Va.	1

Signal Corps

Chief Signal Officer United States Army Washington 25, D.C.	1
Commanding General US Army Combat Surveillance Agency 1124 N. Highland Street Arlington, Virginia	1
Commanding General US Army Electronic Proving Groun. Fort Huachuca, Arizona Attn: Library	1
Commanding General US Army Signal R&D Laboratories Fort Monmouth, New Jersey	1
Commanding Officer US Army White Sands Signal Agency White Sands Proving Ground New Mexico	1
XS, USASRDL Evans Area Belmar, New Jersey	1

Transportation Corps

The Chief of Transportation United States Army Washington 25, D.C.	1
Commanding Officer US Army Transportation Research Command Fort Eustis, Virginia	1
President US Army Transportation Board Fort Eustis, Virginia	1
Commanding General US Army Transportation Materiel Command St. Louis 66 Missouri	1
Commanding Officer US Army Transportation Combat Development Agency Fort Eustis, Virginia	1

Other Army Agencies

Army War College, Library-B-205 Carlisle Barracks Carlisle, Pennsylvania	1
Command and General Staff College Fort Leavenworth, Kansas Attn: Library	1
Commanding General US Continental Army Command Fort Monroe, Virginia	1
US Continental Army Command US Combat Development Experimentation Center Fort Ord, California	1
Office of Special Weapons Development Fort Bliss, Texas	1
Combat Operations Research Group Headquarters, Continental Army Command Fort Monroe, Virginia	1
President US Army Air Defense Board Fort Bliss, Texas	1
Combat Developments Office US Army Infantry School Fort Benning, Georgia	1
Superintendent United States Military Academy West Point, New York	1

DEPARTMENT OF THE NAVY

Chief of Naval Operations Washington 25, D.C.	1
Chief of Naval Research Office of Naval Research (Code 407) Washington 25, D.C.	1
Director US Naval Research Laboratories Washington 25, D.C.	1
Director Weapons Systems Analysis Division Bureau of Naval Weapons Washington 25, D.C.	1

DEPARTMENT OF THE AIR FORCE

Headquarters US Air Force (AFCOA) Washington 25, D.C.	1
Commander Air Force Systems Command Andrews Air Force Base Washington 25, D.C.	1

DEPARTMENT OF THE AIR FORCE (Continued)

Commander 1
Office of Aero Space Research
Building T-D
Washington 25, D.C.

Director 1
Air University Library
Maxwell Air Force Base, Alabama
Attn: AUL-8641

Commanding Officer 1
Wright Patterson Air Force Base, Ohio

OTHER GOVERNMENT AGENCIES

Atomic Energy Commission 1
1901 Constitution Avenue, N.W.
Washington 25, D.C.

National Aeronautics and Space Agency 1
Washington 25, D.C.

National Bureau of Standards 1
Connecticut Avenue & Van Ness Streets, N.W.
Washington 25, D.C.

Human Performance Requirements 1
NASA, Ames Research Center
Moffett Field, California

Director 1
Science Information Exchange
Smithsonian Institution
Washington 25, D.C.

National Bureau of Standards 1
Boulder, Colorado

National Institutes of Health 1
Bethesda, Maryland
Attn: Director

National Science Foundation 1
Washington 25, D.C.

The Library of Congress 1
Washington 25, D.C.

OTHER AGENCIES

British Joint Services Mission 1
Thru: Foreign Liaison Office
Assistant Chief of Staff, Intelligence
Department of the Army
Washington 25, D.C.

Canadian Joint Staff 1
Thru: Foreign Liaison Office
Assistant Chief of Staff, Intelligence
Department of the Army
Washington 25, D.C.

OTHER AGENCIES (Continued)

Human Resources Research Office 1
The George Washington University
2013 G Street, N.W.
Washington 7, D.C.

National Academy of Sciences 1
2101 Constitution Avenue, N.W.
Washington, D.C.

Research Analysis Corporation 1
6935 Arlington Road
Bethesda, Maryland

Special Operations Research Office 1
American University
4501 Massachusetts Avenue, N.W.
Washington 16, D.C.

Mathematics Research Center, US Army 1
University of Wisconsin
Madison, Wisconsin

UNCLASSIFIED

UNCLASSIFIED